

Development and Validation of the German Version of the Large Language Model Dependency Scale

Magnus Liebherr

`magnus.liebherr@uni-due.de`

University Duisburg-Essen

Ala Yankouskaya

Bournemouth University

Sameha AlShakhsi

Hamad Bin Khalifa University

Christian Montag

University of Macau

Raian Ali

Hamad Bin Khalifa University

Research Article

Keywords: Large Language Models, Dependency, Relational Dependency, Instrumental Dependency, GenAI

Posted Date: March 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8997779/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Discover Artificial Intelligence on July 28th, 2026. See the published version at <https://doi.org/10.1007/s44163-026-01781-4>.

Development and Validation of the German Version of the Large Language Model Dependency Scale

Magnus Liebherr¹, Ala Yankouskaya², Sameha AlShakhsi³, Christian Montag⁴,
Raian Ali³

1 Faculty of Engineering, University Duisburg-Essen, Duisburg, Germany

`magnus.liebherr@uni-due.de`

2 Department of Psychology, Bournemouth University, UK

`ayankouskaya@bournemouth.ac.uk`

3 College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

`salshakhsi2@hbku.edu.qa, raali2@hbku.edu.qa`

4 Centre for Cognitive and Brain Sciences, Institute of Collaborative Innovation, University of Macau,
Macau, SAR, China

`cmontag@um.edu.mo`

Abstract

As large language models (LLMs) become increasingly embedded in everyday work, education, and decision-making, concerns have emerged regarding users' reliance on these systems and the psychological mechanisms underlying such reliance. In German-speaking contexts, public discourse on generative AI is strongly shaped by issues of data protection, transparency, autonomy, and regulatory oversight, yet empirically validated instruments to assess dependency on LLMs are lacking. Addressing this gap, the present study translated and validated the German version of the Large Language Model Dependency Scale (LLM-D12-DE). A sample of 402 German-speaking active LLM users completed the German LLM-D12. Confirmatory factor analysis supported the hypothesized two-factor structure comprising Instrumental Dependency and Relationship Dependency. Both dimensions demonstrate excellent internal consistency, strong composite reliability, and satisfactory convergent validity. Scalar measurement invariance across gender was established, allowing meaningful latent mean comparisons. Network analysis revealed two coherent but distinct item clusters corresponding to instrumental reliance and relational attachment, with central items reflecting decision-related reliance and perceived reduction of loneliness. External validation analyses showed that both dependency dimensions were positively associated with AI acceptance attitudes, perceived trustworthiness of LLMs, and internet addiction severity. Relationship Dependency was additionally associated with lower need for cognition. Usage characteristics, particularly higher frequency of LLM use and multimodal interaction, emerged as robust predictors of dependency levels. Overall, the findings demonstrate that the LLM-D12-DE is a psychometrically sound and contextually appropriate instrument for assessing dependency on LLMs in German-speaking populations. The scale enables systematic research on cognitive and relational aspects of human-LLM interaction and provides an empirical basis for educational, clinical, and regulatory discussions on responsible AI use.

Keywords: Large Language Models, Dependency, Relational Dependency, Instrumental Dependency, GenAI

1 Introduction

Large Language Models (LLMs) such as ChatGPT, Gemini, or DeepSeek have rapidly become embedded in both professional and private life, profoundly shaping how people search for information, solve problems, and make decisions (Dwivedi et al., 2023; Noy & Zhang, 2023). Their widespread adoption has created new opportunities for productivity and learning but also raised concerns about the extent to which individuals may develop strong forms of dependence on these systems (Yankouskaya, Liebherr, et al., 2025). This duality highlights the necessity of systematic research into the psychological mechanisms underlying human-LLM interaction, particularly regarding the balance between functional benefits and potential risks.

LLMs differ fundamentally from other types of technology. Whereas traditional digital systems are typically rule-based, deterministic, and optimized for narrowly defined tasks, LLMs compute token sequences by estimating probability distributions trained on massive datasets of natural language. This design enables a remarkable degree of generalization and flexible application across domains, ranging from scientific writing support to customer service. However, their open-endedness also makes them more difficult to control and predict, raising questions about trust, transparency, and appropriate reliance. From a psychological perspective, LLMs can be conceptualized along two interrelated dimensions. First, they provide cognitive support by acting as tools for cognitive offloading (Risko & Gilbert, 2016). By externalizing memory, reasoning, and text production, LLMs conserve scarce executive resources and may induce states of heightened productivity or even flow. However, this comes at a cost: frequent reliance may encourage automation bias and diminish independent critical thinking (Dergaa et al., 2024). Their capacity to deliver instant, tailored responses reinforces habits of delegation and gratification, potentially leading to compulsive use patterns that parallel other forms of behavioral dependency (Yankouskaya, Liebherr, et al., 2025). Second, LLMs foster emotional engagement by simulating reciprocal conversation, social presence, and even affective validation. Unlike earlier tools, they may act as social actors rather than mere instruments (Sundar & Lee, 2022). Through personalization and continuity, LLMs generate a sense of self-specificity, interactions that feel uniquely relevant to one's identity and needs (Giubilini et al., 2024; Zhang et al., 2024; Zhu et al., 2024). This dynamic enhances the likelihood of forming parasocial bonds or even deeper pseudosocial connections (Horton & Richard Wohl, 1956; Skjuve et al., 2021), which resemble real human relationships despite their one-sidedness. Users may thus perceive LLMs as companions, blurring the boundary between functional support and emotional intimacy.

This intersection of instrumental task support and relational engagement distinguishes LLM dependency from related constructs such as social media disorder. As argued by Yankouskaya, Liebherr, et al. (2025), while social media disorder involves human-to-human interaction mediated by technology, LLM dependency occurs in the absence of a sentient agent other than the user. There is no human counterpart involved, which fundamentally alters the psychological dynamics of the dependency

relationship. Despite their transformative potential and these unique characteristics, validated measurement tools to capture LLM dependency remain scarce. Early scales such as the Problematic ChatGPT Use Scale (Yu et al., 2024) or the AI Chatbot Dependence Scale (Zhang et al., 2025) predominantly adapt criteria derived from the Diagnostic and Statistical Manual of Mental Disorders, representing a subset of addiction criteria originally developed for gaming or internet use disorders. While useful as initial attempts, they often fail to account for the dialogic interactivity, anthropomorphic design, and cognitive partnership that define LLMs (Hu et al., 2023; Huang et al., 2024). To address the conceptual and measurement gap identified in their earlier work on whether ChatGPT can be addictive (Yankouskaya, Liebherr, et al., 2025), Yankouskaya, Babiker, et al. (2025) introduced the LLM-D12, a dual-dimensional scale designed to measure both instrumental and relational forms of LLM dependency. The validation study in the UK demonstrated excellent psychometric properties and discriminant validity (Yankouskaya, Babiker, et al., 2025), while an Arabic version confirmed the robustness of the two-factor structure across cultures (AlShakhsi et al., under review).

The need for culturally and linguistically adapted versions of the LLM-D12 extends beyond mere translation. Cross-cultural research has repeatedly shown that technology-related attitudes and behaviors vary systematically across sociocultural contexts. Cultural values shape not only the prevalence rates of behavioral addictions but also how dependency symptoms are expressed, interpreted, and socially evaluated (Kuss et al., 2013; Wartberg et al., 2020). Recent comparative studies have demonstrated that participants from Arab-Islamic regions report consistently higher readiness to delegate tasks to AI systems than Anglo-individualistic populations, while Germanic European cultures occupy an intermediate position that remains domain-sensitive, varying across contexts such as companionship, health advice, or education (Liebherr et al., under review). These findings underscore that both the instrumental (cognitive) and relational (emotional) aspects of LLM dependency may manifest differently across populations, necessitating rigorous validation to ensure conceptual equivalence and enable meaningful cross-cultural comparisons. Translating and validating the LLM-D12 in the German context constitutes a critical next step for several reasons. Germany is currently engaged in intensive debates surrounding the regulation of generative AI, particularly concerning data privacy, transparency requirements, and the protection of individual autonomy under frameworks such as the “General Data Protection Regulation” and the emerging EU AI Act. Public attitudes toward AI in Germany also tend to emphasize caution and ethical consideration; for instance, comparative research shows that German participants were more cautious and reflective in their assessments of AI than Chinese participants, who tended to view AI’s societal benefits more optimistically (Brauner et al., 2024). Understanding LLM dependency in this context is essential for evidence-based policymaking. For instance, if individuals with higher LLM dependency exhibit reduced concern for data privacy or diminished critical evaluation of AI-generated content, this has direct implications for regulatory effectiveness and digital literacy initiatives. Furthermore, the German educational system is grappling with questions of how to integrate LLMs into teaching and learning while preserving pedagogical goals such as independent reasoning,

source evaluation, and academic integrity (Ibeh et al., 2025). Valid measurement of LLM dependency is necessary to investigate whether reliance on these systems affects educational outcomes, cognitive skill development, and students' willingness to engage in effortful learning processes. Without culturally validated instruments, research on these questions remains limited in its generalizability and practical applicability.

Establishing the validity of the LLM-D12 in German-speaking populations will thus provide a foundation for cross-cultural measurement equivalence, contribute to the global mapping of LLM dependency, and inform both scientific understanding and policy discussions on the cognitive and emotional implications of human-AI interaction in culturally specific contexts. Therefore, the present study aimed to translate and psychometrically validate the LLM-D12 for German-speaking populations (LLM-D12-DE).

2 Methods

2.1 Participants and procedure

Data were collected between August and September 2025, using the online platform Prolific (www.prolific.com). The survey was administered via SurveyMonkey (www.surveymonkey.com) and conducted in German. All participants provided digital informed consent prior to the study. The study was approved by the local ethics committee of the University Bournemouth University, UK (62239, 03/03/2025). The study was conducted anonymously, and participants who successfully completed the survey and passed all attention checks received monetary compensation. The final sample comprised 402 participants (mean age = 30.93 years, SD = 6.08). Table 1 presents detailed demographic characteristics of the sample.

Table 1. Demographics

	Categories	N	Proportion
Education	Primary education (elementary)	29	7.21
	Pursuing or completed postgraduate degree (master's, Ph.D., etc.)	99	24.63
	Pursuing or completed undergraduate degree (bachelor's)	140	34.83
	Pursuing or completed vocational or technical education	86	21.39
	Secondary education (high school)	48	11.94
Employment	Full time employment	245	60.95

	Other (please specify)	8	1.99
	Part time employment	48	11.94
	Retired	1	0.25
	Run my own business	8	1.99
	Student	70	17.41
	Unemployed	22	5.47
Gender	Male	205	51.00
	Female	194	48.30
	No response	3	0.70

Initially, prospective participants were screened for their engagement with LLMs. Therefore, individuals were asked about their usage patterns with the LLM they use most frequently, which we termed their "Primary LLM". Participants were eligible for inclusion if they met at least one of the following criteria: (1) they used their Primary LLM on a near-daily basis and reported substantial reliance on it; (2) they used it on a near-daily basis without necessarily reporting high reliance; or (3) they reported substantial reliance despite less frequent usage. Individuals who neither used their Primary LLM regularly nor reported meaningful reliance were excluded from participation. This ensured that the scale was administered to individuals for whom the construct of LLM dependency is most relevant. Additional eligibility requirements included being aged 18 to 40 years of age and self-identifying as a resident of a German-speaking region with German as their primary or fluent language.

2.2 Translation process

The German version of the LLM-D12 was developed following established guidelines for cross-cultural adaptation of psychometric instruments. We employed a systematic back-translation procedure to ensure both linguistic accuracy and conceptual equivalence [Ref]. First, two independent bilingual translators, both native German speakers with fluency in English and expertise in psychology and human-computer interaction, independently translated the original English version into German. Second, discrepancies between the two translations were discussed and resolved through consensus, with particular attention to preserving the theoretical meaning of each item while ensuring natural German phrasing. Third, a different bilingual translator, not previously involved in the forward translation and unfamiliar with the original scale, independently translated the reconciled German version back into English. Finally, the senior author compared the back-translated English version with the original LLM-D12 to verify semantic equivalence and identify any unintended shifts in meaning.

2.3 Measures

2.3.1 LLM Dependency Scale (LLM-D12-DE)

The German version of the LLM-D12 consists of 12 items assessing two theoretically distinct but empirically related dimensions of dependency on Large Language Models. The scale was originally developed by Yankouskaya, Babiker, et al. (2025), which conceptualizes LLM dependency as arising from both instrumental and relational forms of engagement.

Instrumental Dependency (6 items; item codes: ID1-ID6) captures users' habitual reliance on LLMs for cognitive support, including memory assistance, decision-making processes, task completion, and problem-solving. This dimension reflects patterns of cognitive offloading and task absorption, whereby users delegate mental effort to the LLM and experience heightened reward or flow states during interaction. Example items include: "Without it, I feel less confident when making decisions" (*“Wenn ich Entscheidungen ohne mein primäres LLM treffe, fühle ich mich unsicher”*) and "It is my go-to for assistance in decision-making" (*“Mein primäres LLM ist meine erste Anlaufstelle, wenn ich Hilfe bei der Entscheidungsfindung brauche”*). Scores are computed as the sum of the six items, ranging from 6 to 36, with higher scores indicating greater instrumental dependency.

Relationship Dependency (6 items; item codes: RD1-RD6) assesses affective and social attachment to LLMs, including the attribution of human-like qualities (anthropomorphism), formation of parasocial bonds, emotional disclosure, and substitution of human interaction. This dimension is theoretically grounded in research on parasocial relationships and social compensatory use of technology. Example items include: "It helps me feel less alone when I need to talk to someone" (*“Mein primäres LLM hilft mir, mich weniger allein zu fühlen, wenn ich mit jemandem reden muss”*) and "I interact with it as if it were a genuine companion" (*“Ich interagiere mit meinem primären LLM so, als wäre es eine reale Bezugsperson”*). One item (RD5: "I use it solely as a tool, not to express my feelings or expect it to understand me"/*“Ich benutze mein primäres LLM nur als Hilfsmittel, nicht um meine Gefühle auszudrücken oder zu erwarten, dass es mich versteht”*) is reverse-scored. Scores are computed as the sum of the six items (after reverse-coding), ranging from 6 to 36, with higher scores indicating greater relationship dependency.

Total LLM Dependency can optionally be calculated as the sum of both subscales, ranging from 12 to 72. All items were rated on a 6-point Likert scale: 1 = *Stimme überhaupt nicht zu* (Strongly Disagree), 2 = *Stimme nicht zu* (Disagree), 3 = *Stimme eher nicht zu* (Somewhat Disagree), 4 = *Stimme eher zu* (Somewhat Agree), 5 = *Stimme zu* (Agree), 6 = *Stimme voll und ganz zu* (Strongly Agree). The absence of a neutral midpoint was intended to encourage respondents to indicate a directional stance rather than defaulting to ambivalence. A comprehensive list of all LLM-D12-DE items in both German and English, along with item codes and detailed scoring instructions, is provided in Appendix 1.

2.3.2 Internet Addiction (IA)

Internet addiction was measured using the 7-item short form of the Internet Addiction Test (IAT-7) (Valenti et al., 2025). Participants rated each item on a 5-point Likert scale from 1 (Never) to 5 (Always). Total scores range from 7 to 35, with higher scores indicating greater internet addiction severity. The scale showed good internal consistency in the present sample (Cronbach's $\alpha = .799$).

2.3.3 Attitudes Toward AI (ATAI)

Attitudes toward AI were measured using the ATAI scale (Sindermann et al., 2021), which comprises two subscales:

1. ATAI Acceptance (2 items) assesses positive attitudes toward AI adoption and integration. Items were rated on an 11-point Likert scale from 0 (Strongly Disagree) to 10 (Strongly Agree). Total scores range from 0 to 20, with higher scores reflecting more positive attitudes.
2. ATAI Fear (3 items) assesses anxiety and concerns about AI development and deployment. Items were rated on the same 11-point scale, with total scores ranging from 0 to 30. Higher scores indicate greater fear or apprehension toward AI.

Internal consistency in the current study was acceptable for both subscales (Cronbach's $\alpha = .832$ for Acceptance; $\alpha = .807$ for Fear).

2.3.4 Need for Cognition (NFC)

Need for cognition was assessed using the 6-item short version of the NFC scale (Lins de Holanda Coelho et al., 2020), which measures individuals' intrinsic motivation to engage in effortful cognitive activities. Items were rated on a 5-point Likert scale ranging from 1 (Extremely uncharacteristic of me) to 5 (Extremely characteristic of me). Total scores range from 6 to 30, with higher scores indicating greater need for cognition. The scale demonstrated acceptable internal consistency in this sample (Cronbach's $\alpha = .809$).

2.3.5 Perceived Trustworthiness of primary LLM (PT-LLM)

Perceived trustworthiness of the participant's Primary LLM was assessed using an 8-item scale adapted from an established multidimensional framework of AI trustworthiness (Huang et al., 2024; Yankouskaya, Babiker, et al., 2025). The scale captures eight distinct facets: accuracy and truthfulness, safety, fairness, robustness, privacy protection, ethical behavior, transparency, and accountability. Each item was rated on an 11-point Likert scale ranging from 0 (Not at all) to 10 (Completely). An example item is: "I trust my primary LLM to provide factually accurate and reliable information". Scores were summed to yield a total trustworthiness score ranging from 0 to 80, with higher scores indicating greater perceived trustworthiness. Internal consistency in the present sample was excellent (Cronbach's $\alpha = .933$).

2.3.6 The use of LLMs

Participants provided information about their Primary LLM usage patterns, including: (1) primary device used to access the LLM (mobile, PC and other devices); (2) frequency of use (categorical: Once daily; 1–2 times per week; 1–3 times per month; Less than once per month); (3) interaction method (text, voice, text and voice); and (4) typical location of use (stationary such as home or office, both stationary and on the move).

2.3.7 Attention Checks

To ensure data quality, the survey included attention check items strategically embedded throughout the questionnaire. These items were designed to match the structure, length, and tone of substantive scale items, making them less conspicuous while still enabling identification of inattentive or careless responding. Participants who failed two or more attention checks were excluded from analysis.

2.4 Data Analysis

2.4.1 Monotonicity check

Monotonicity was examined using the monotonicity check implemented in the *mokken* package in R, which evaluates whether item response probabilities show non-decreasing patterns across increasing total scores. Monotonicity analyses indicated that none of the Instrumental dependency items departed from the expected non-decreasing pattern across the total score range. Response probabilities increased consistently with higher latent scores for all instrumental dependency items, and no violations were identified at the item-level total. For the Relationship subscale, the results indicated a single departure from monotonicity for one reverse-coded item (*solely_as_a_tool*), reflected in a local decrease of 0.10 in response probability at one point on the score continuum. This departure was isolated to a single score transition and was not accompanied by additional decreases across the remaining score range. All other items showed non-decreasing response probabilities throughout (see Supplementary Material 1 for monotonicity plots).

2.4.2 Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was conducted to test the hypothesised two-factor structure of the LLM-D-12 scale, comprising Instrumental dependency and Relationship dependency. All items were measured on six-point Likert-type response scales and were therefore treated as ordinal indicators. Models were estimated in *lavaan* using robust weighted least squares estimation (WLSMV), which is appropriate for ordinal data and is based on the polychoric correlation structure among items. Latent factor variances were fixed to unity to identify the model (*std.lv* specification), allowing factor loadings to be interpreted on a standardised metric. Model fit was evaluated using multiple complementary indices, including the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA) with 90% confidence intervals, and the standardised root mean square residual (SRMR). Standardised factor loadings, residual variances, and the latent factor correlation were

examined to assess the adequacy of the measurement model. Modification indices were inspected descriptively to identify potential areas of local misfit, but no post-hoc model modifications were implemented.

2.4.3 Reliability and convergent validity

Internal consistency and convergent validity were examined using reliability indices appropriate for ordinal indicators. Ordinal reliability was estimated for each subscale using ordinal alpha and omega total, computed from polychoric correlation matrices, given the six-point Likert response format. In addition, CFA-based composite reliability (CR) and average variance extracted (AVE) were calculated from the standardised factor loadings and item-level explained variance obtained from the confirmatory factor analysis. Composite reliability was used to evaluate internal consistency at the latent level, while AVE was used to assess the extent to which each latent factor accounted for variance in its observed indicators. All reliability and validity indices were computed separately for the Instrumental and Relationship dependency dimensions. Item-total correlations were examined to assess the degree to which each item was associated with its corresponding subscale score.

2.4.4 Measurement invariance analysis

Measurement invariance of the two-factor model was examined across gender using multi-group confirmatory factor analysis. Gender was selected as the grouping variable to evaluate whether the measurement properties of the Instrumental and Relationship dependency factors were equivalent for male and female participants, thereby ensuring that observed group comparisons reflect substantive differences rather than measurement artefacts. Of the total sample ($N = 402$), participants who identified as male ($n = 205$) or female ($n = 194$) were retained for analysis. Three participants who did not specify their gender were excluded prior to model estimation.

Models were estimated using the weighted least squares mean- and variance-adjusted estimator with thresholds specified for all indicators. A hierarchical sequence of increasingly restrictive models was tested. First, a configural invariance model was estimated to establish whether the same factor structure held across gender groups. Second, metric invariance was tested by constraining all factor loadings to equality across groups, assessing whether items related to the latent constructs in the same way for males and females. Third, scalar invariance was evaluated by additionally constraining item thresholds to equality, allowing for comparisons of latent factor means across groups. Model fit was evaluated using the comparative fit index, Tucker–Lewis index, root mean square error of approximation with 90% confidence intervals, and the standardised root mean square residual. Changes in model fit across nested models were used to determine support for measurement invariance.

2.4.5 Network analysis

To complement the latent variable analyses, a network analytic approach was used to examine the structure of associations among the 12 items of the German LLM-D12. Network analysis allows the

modelling of conditional relationships between observed variables by estimating partial correlations, thereby identifying direct associations between items while controlling for all others in the network.

Responses were treated as ordinal and polychoric correlations were computed to account for the ordered categorical response format. A sparse partial correlation network was estimated using the graphical least absolute shrinkage and selection operator (glasso) with Extended Bayesian Information Criterion model selection. This regularisation procedure shrinks small associations towards zero and retains only the most robust edges, reducing the likelihood of spurious connections and improving interpretability. Node importance was quantified using strength and expected influence. Strength reflects the sum of absolute edge weights connected to a node, indicating the overall connectedness of an item within the network, whereas expected influence incorporates the direction of associations and is therefore sensitive to the presence of negative edges. Edge weights represent regularised partial correlations, with positive values indicating direct positive associations and negative values indicating inverse associations between items.

2.4.6 External validation

To examine the external validity of the LLM-D12-DE, two separate multiple regression models were estimated. The aim of this analysis was to quantify the extent to which each LLM-D12 subscale was associated with theoretically relevant external constructs after adjusting for shared variance among predictors. In the first model, the instrumental LLM dependency score served as the dependent variable; in the second model, the relationship score was the dependent variable. In both models, perceived trustworthiness of primary LLM (PT-LLM-8), attitudes toward artificial intelligence (ATAI acceptance and ATAI fear), need for cognition (NFC), and Internet addiction total score were included simultaneously as predictors.

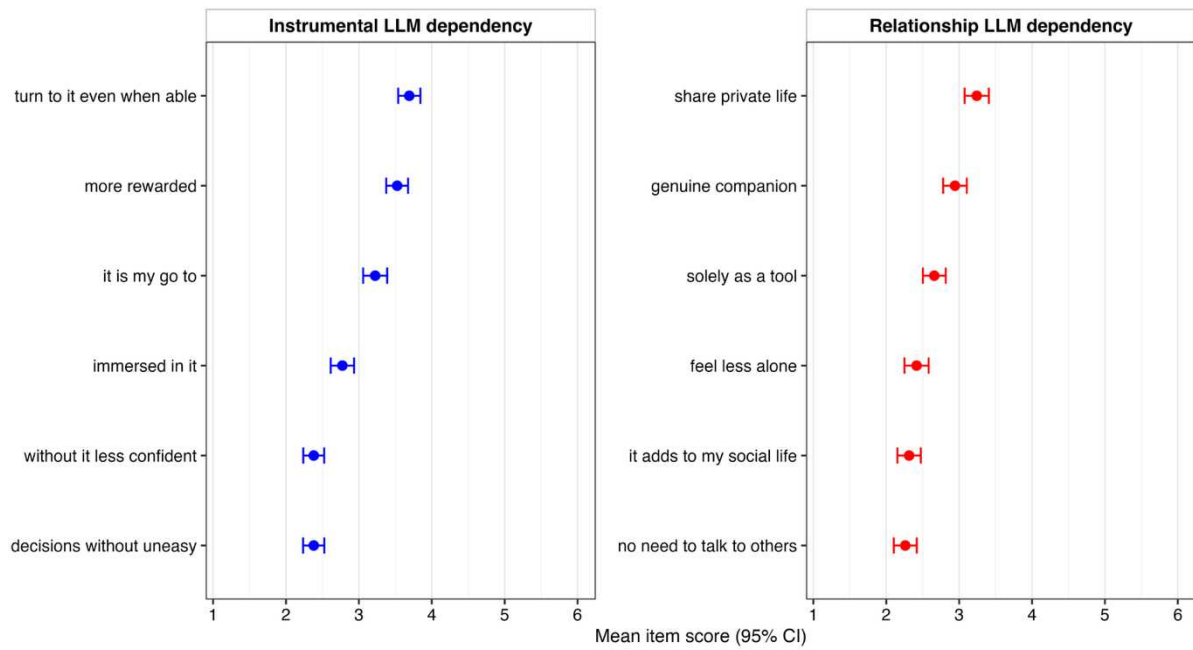
Robust multiple regression was used to reduce sensitivity to violations of normality, heteroscedasticity, and influential observations. Models were estimated using MM-estimation, which combines high breakdown point estimation with efficient asymptotic properties. Regression coefficients, robust standard errors, and 95% confidence intervals were computed for all predictors. Model fit was summarised using robust coefficients of determination (R^2 and adjusted R^2).

3 Results

3.1 Descriptive statistics

The mean score for Instrumental LLM dependency was 2.99 (SD = 1.32), and for Relationship LLM dependency was 2.63 (SD = 1.34). Descriptive statistics for individual items are presented in Figure 1.

Figure 1. Mean item scores (95% confidence intervals) for Instrumental and Relationship LLM dependency



3.2 Confirmatory Factor Analysis

The latent factors were found to be strongly and positively correlated. Specifically, the standardised latent correlation between Instrumental and Relationship dependency was 0.863, with a standard error of 0.015 and a corresponding z-value of 57.05, $p < .001$.

The chi-square test of model fit was significant, $\chi^2(53) = 115.36$, $p < .001$, which is expected in confirmatory factor analysis because the chi-square statistic increases with sample size (Bentler & Bonett, 1980). The hypothesised two-factor CFA model provided an excellent fit to the data with a comparative fit index of 0.99 and a Tucker-Lewis index of 0.99. The RMSEA was 0.054, with a 90% confidence interval ranging from 0.041 to 0.068, which falls below the commonly used 0.07 threshold for acceptable model fit (Steiger, 2007). The SRMR was 0.040, further supporting the adequacy of the measurement model.

Standardised factor loadings for the two-factor confirmatory factor analysis are reported in Table 2. All items loaded positively and significantly on their latent factors, supporting the hypothesised measurement structure. Loadings were generally strong across both dimensions. One item on the Relationship dependency factor (*solely_as_a_tool_r*) showed the weakest loading relative to the other indicators, although it still contributed substantially to the factor.

Squared multiple correlations (R^2) indicated that the Instrumental dependency items were generally well explained by their latent factor, with values ranging from .59 to .85 (Table 2). Most instrumental items showed high explained variance ($R^2 \geq .69$), indicating that a substantial proportion of item variance was accounted for by the underlying construct. For the Relationship dependency factor, explained variance was very high for several core indicators, including *feel_less_alone*, *it_adds_to_my_social_life*, and *no_need_to_talk_to_others*. *Genuine_companion* also demonstrated

strong explained variance, while *share_private_life* showed a moderate level of variance explained. Consistent with its comparatively weaker loading, *solely_as_a_tool_r* exhibited substantially lower explained variance, indicating that this item was less strongly accounted for by the relationship dependency factor than the other indicators.

Table 2. Standardised factor loadings for the two-factor confirmatory factor analysis of the LLM-D-12 scale

Factor	Item code	λ	SE	z	p-value	R ²
Instrumental LLM	without_it_less_confident	0.867	0.017	50.332	0.000	0.752
Dependency	immersed_in_it	0.832	0.019	44.632	0.000	0.691
	more_rewarded	0.849	0.016	51.936	0.000	0.721
	turn_to_it_even_when_able	0.770	0.023	33.943	0.000	0.593
	it_is_my_go_to	0.905	0.013	68.067	0.000	0.819
	decisions_without_uneasy	0.922	0.013	70.010	0.000	0.850
Relationship LLM	share_private_life	0.744	0.027	27.953	0.000	0.553
Dependency	genuine_companion	0.861	0.016	52.636	0.000	0.741
	feel_less_alone	0.953	0.008	115.783	0.000	0.907
	it_adds_to_my_social_life	0.950	0.009	106.184	0.000	0.903
	solely_as_a_tool_r	0.418	0.042	9.955	0.000	0.174
	no_need_to_talk_to_others	0.947	0.009	105.742	0.000	0.896

Note. Λ – standardised factor loading, representing the strength of the relationship between the item and its latent factor; SE – the standard error of the factor loading estimate, reflecting the sampling variability of the loading, z – Wald test testing whether the estimate differs from zero (if p-value < 0.05, the null hypothesis can be rejected), R² - represents the proportion of variance in an observed indicator that is explained by its latent factor

Residual correlations ranged from -0.11 to 0.12. Using a commonly applied criterion for local dependence ($|\text{residual}| > 0.1$), no item pairs exceeded the threshold (Knekta et al., 2019).

Reliability and convergent validity

Internal consistency was high for both dimensions of the LLM-D12 when estimated using ordinal reliability coefficients based on polychoric correlations. The Instrumental dependency subscale demonstrated excellent reliability (ordinal $\alpha = .94$; $\omega_t = .94$; 6 items). Similarly, the Relationship dependency subscale showed strong internal consistency (ordinal $\alpha = .92$; $\omega_t = .93$; 6 items). The close

correspondence between ordinal alpha and omega total across both subscales indicates a coherent item structure with minimal evidence of multidimensionality within each factor.

CFA-based composite reliability was high for both latent dimensions. The Instrumental dependency factor showed excellent composite reliability ($CR = .94$), while the Relationship dependency factor also demonstrated strong reliability ($CR = .93$). These values indicate a high degree of internal consistency at the latent level. Convergent validity was supported for both factors. The average variance extracted (AVE) exceeded conventional benchmarks for each dimension, with values of .74 for Instrumental dependency and .70 for Relationship dependency, indicating that the latent constructs accounted for a substantial proportion of variance in their respective indicators. Correspondingly, the square roots of AVE were high for both factors (Instrumental: $\sqrt{AVE} = .86$; Relationship: $\sqrt{AVE} = .83$), reflecting strong shared variance between each construct and its items.

Corrected item-total correlations indicated that all items were positively associated with their respective subscale scores. For the Instrumental dependency subscale, item-total correlations were consistently high, ranging from .72 to .86, with all six items showing strong associations with the overall subscale score. For the Relationship dependency subscale, item-total correlations ranged from .43 to .88. Most items showed strong associations with the subscale score, while one item (*solely_as_a_tool_r*) demonstrated a comparatively weaker, but still positive, association relative to the other indicators.

Overall, these results provide robust evidence for both latent reliability and convergent validity of the two LLM dependency dimensions.

Measurement invariance

Multi-group confirmatory factor analysis supported measurement invariance of the two-factor model across groups. The configural model demonstrated good fit ($\chi^2(106) = 146.75, p = .005$; CFI = .999; TLI = .999; RMSEA = .044, 90% CI [.025, .060]; SRMR = .043), indicating that the same factor structure held across groups. Constraining all factor loadings to equality (metric invariance) resulted in similarly good fit ($\chi^2(116) = 177.19, p < .001$; CFI = .999; TLI = .999; RMSEA = .052, 90% CI [.036, .066]; SRMR = .048), with minimal changes relative to the configural model ($\Delta CFI = -.0004$, $\Delta RMSEA = +.0075$, $\Delta SRMR = +.0045$). Further constraining item thresholds to equality (scalar invariance) did not degrade model fit ($\chi^2(162) = 221.99, p = .001$; CFI = .999; TLI = .999; RMSEA = .043, 90% CI [.028, .057]; SRMR = .044), with negligible changes relative to the metric model ($\Delta CFI = +.00002$, $\Delta RMSEA = -.0084$, $\Delta SRMR = -.0035$). These results indicate that both metric and scalar invariance were supported.

Under the metric invariance model, all constrained factor loadings were statistically significant ($ps < .001$). Loadings on the Instrumental factor were uniformly high, ranging from .80 to .91, while loadings on the Relationship factor ranged from .45 to .96. The item *solely_as_a_tool_r* showed the weakest loading on the Relationship factor ($\lambda \approx .45$), whereas *feel_less_alone*,

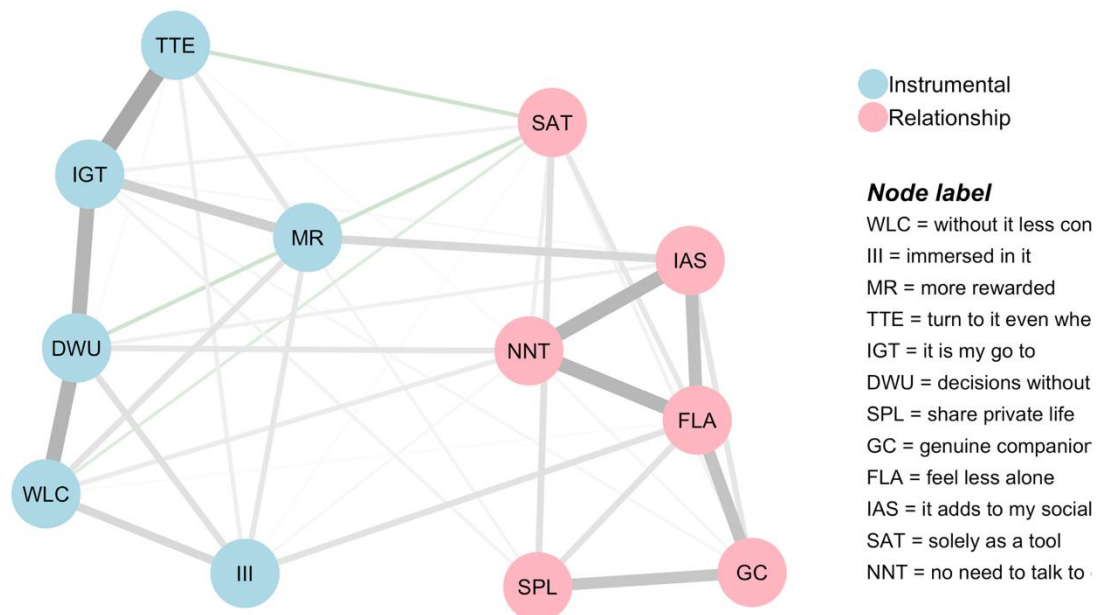
it_adds_to_my_social_life, and *no_need_to_talk_to_others* showed the strongest associations ($\lambda_s \geq .94$). Although standardised loadings differed modestly between groups, unstandardised loadings were identical by constraint, consistent with metric invariance.

The latent correlation between Instrumental and Relationship dependency was large in both groups (group 1: $r = .88$, 95% CI [.85, .92], $p < .001$; group 2: $r = .62$, 95% CI [.40, .84], $p < .001$). With scalar invariance established, latent mean comparisons indicated a lower mean on Instrumental dependency in the second group relative to the reference group (-0.46 , $SE = 0.11$, $p < .001$), whereas the latent mean difference for Relationship dependency was not statistically significant (-0.20 , $SE = 0.11$, $p = .065$).

Network analysis

Across the network, centrality estimates indicated heterogeneity in node importance (Figure 2).

Figure 2. Network structure of the German LLM-D12



Note. Edges represent regularised partial correlations between items, estimated using EBICglasso. Edge thickness reflects the magnitude of the association. Grey edges indicate positive partial correlations, whereas green edges indicate negative partial correlations.

Within the Instrumental LLM dependency subscale, *it_is_my_go_to* showed the highest centrality (Strength = 1.26; Expected Influence = 1.26), followed by *decisions_without_uneasiness* (Strength = 1.17; Expected Influence = 1.01). These nodes occupied the most influential positions among instrumental indicators, reflecting strong direct connections with multiple other items. *Without_it_less_confident* and *more_rewarded* also demonstrated relatively high centrality (Strengths = 0.92 and 0.86, respectively), whereas *immersed_in_it* and *turn_to_it_even_when_able* showed more moderate centrality.

Examination of edge weights showed that associations were predominantly positive within the Instrumental LLM dependency subscale. The strongest partial correlations were observed between *turn_to_it_even_when_able* and *it_is_my_go_to* (weight = 0.43), between *without_it_less_confident* and *decisions_without_uneasy* (0.37), and between *it_is_my_go_to* and *decisions_without_uneasy* (0.37). These connections suggest a tightly interconnected instrumental cluster centred on habitual use and decision reliance.

Within the Relationship LLM dependency subscale, *feel_less_alone* emerged as the most central node overall (Strength = 1.38; Expected Influence = 1.38). This was followed by *it_adds_to_my_social_life* (Strength = 1.09; Expected Influence = 1.09) and *no_need_to_talk_to_others* (Strength = 1.10; Expected Influence = 1.08). *genuine_companion* showed moderate centrality (Strength = 0.85), whereas *share_private_life* was comparatively less central. *Solely_as_a_tool_r* indicated a distinctive profile: although its Strength was moderate (0.73), its Expected Influence was substantially lower (0.20), reflecting the presence of negative associations that offset its positive connections. The strongest edges were observed between *it_adds_to_my_social_life* and *no_need_to_talk_to_others* (0.36), between *feel_less_alone* and *no_need_to_talk_to_others* (0.35), between *feel_less_alone* and *it_adds_to_my_social_life* (0.32), and between *share_private_life* and *genuine_companion* (0.28). These edges indicate a coherent relational subnetwork linking social supplementation, emotional connection, and reduced reliance on human interaction.

Cross-subscale connections were present but generally weaker. Notable positive links included *more_rewarded* with *it_adds_to_my_social_life* (0.19) and *decisions_without_uneasy* with *it_adds_to_my_social_life* (0.07), suggesting limited but meaningful bridges between instrumental reward processes and relational experiences with individual's primary LLM. Negative edges were sparse and primarily involved *solely_as_a_tool_r*, which showed small negative associations with several instrumental items, including *without_it_less_confident* (-0.05), *immersed_in_it* (-0.01), *more_rewarded* (-0.04), *turn_to_it_even_when_able* (-0.08), and *decisions_without_uneasy* (-0.08).

Overall, the network structure was characterised by two densely connected clusters corresponding to Instrumental and Relationship LLM dependency.

External validation

External validity of the LLM-D12 was examined using two separate robust multiple regression models predicting instrumental and relationship-based LLM dependency. Both models were estimated on the full sample (N = 402) and included the same set of predictors: perceived trustworthiness of LLM (PT-LLM-8), ATAI acceptance, ATAI fear, need for cognition, and internet addiction score.

The model predicting instrumental LLM dependency explained a substantial proportion of variance (adjusted $R^2 = .44$). Higher instrumental dependency was significantly associated with greater perceived trustworthiness of LLM, stronger AI acceptance attitude, and higher internet addiction scores. AI fear

attitudes and need for cognition were not statistically associated with instrumental LLM dependency (Table 3).

The model predicting relationship LLM dependency also demonstrated moderate explanatory power (adjusted $R^2 = .35$). Relationship dependency was positively associated with perceived trustworthiness of LLM, AI acceptance attitudes, and internet addiction severity. In contrast to the instrumental model, need for cognition showed a significant negative association with relationship LLM dependency, whereas AI fear attitudes were not significantly associated.

Table 3. Robust multiple regression models predicting relationship-based and instrumental LLM dependency

Predictor	Estimate	SE	t	p	95% CI
<i>Instrumental LLM dependency, Model fit: $R^2 = .45$, adjusted $R^2 = .44$</i>					
Intercept	-3.94	2.62	-1.50	.134	[-9.10, 1.21]
PT-LLMs	0.12	0.03	4.61	< .001	[0.07, 0.17]
ATAI acceptance	0.78	0.12	6.64	< .001	[0.55, 1.01]
ATAI fear	-0.04	0.06	-0.73	.468	[-0.16, 0.07]
Need for cognition	-0.04	0.08	-0.53	.593	[-0.19, 0.11]
Internet addiction	0.36	0.06	6.08	< .001	[0.24, 0.47]
<i>Relationship LLM Dependency, Model fit: $R^2 = .35$, adjusted $R^2 = .35$</i>					
Intercept	1.08	2.96	0.36	.716	[-4.74, 6.90]
PT-LLMs	0.06	0.03	2.15	.032	[0.01, 0.12]
ATAI acceptance	0.75	0.14	5.47	< .001	[0.48, 1.03]
ATAI fear	-0.08	0.07	-1.17	.242	[-0.22, 0.06]
Need for cognition	-0.21	0.08	-2.45	.015	[-0.38, -0.0]
Internet addiction	0.37	0.06	6.03	< .001	[0.25, 0.49]

Across both models, AI acceptance attitudes, PT-LLMs and internet addiction emerged as consistent positive correlates of LLM dependency, while AI fear attitudes did not show independent associations when other predictors were controlled. To test whether the associations between external validation variables and LLM dependency differed between instrumental and relationship LLM dependency, Wald

tests were conducted on the outcome-by-predictor interaction terms in the stacked robust regression model. The difference in effects between instrumental and relationship dependency was not significant for PT-LLMs ($F(1, 792) = 2.18, p = .140$), ATAI acceptance ($F(1, 792) = 0.01, p = .904$), ATAI fear ($F(1, 792) = 0.20, p = .657$), need for cognition ($F(1, 792) = 2.23, p = .136$), or internet addiction ($F(1, 792) = 0.02, p = .883$). These results indicate that, although several predictors were robustly associated with LLM dependency overall, their relative strengths did not differ significantly between instrumental and relationship-based dependency.

Additional analysis

Table 4 presents descriptive statistics for instrumental and relationship LLM dependency scores across categories of primary LLM usage.

Table 4. Descriptive statistics for instrumental and relationship LLM dependency by primary LLM usage characteristics

LLM usage	Category	N	Instrumental LLM dependency		Relationship LLM dependency	
			Mean	SD	Mean	SD
Primary LLM device	Mobile	215	18.702	7.729	16.865	8.253
	PC and other devices	187	17.139	8.115	14.647	7.702
Frequency of use primary LLM	Less than 1-2 times per week	185	13.395	5.776	12.405	6.160
	3-6 times per week	61	17.918	6.714	14.475	7.164
	Multiple times daily	156	23.429	7.144	20.429	8.214
Interaction method with primary LLM	Text	276	15.866	6.807	13.841	6.976
	Text and Voice	126	22.728	8.200	20.272	8.585
Location of primary LLM use	Both - Stationary and On the move	212	18.991	7.942	16.896	7.996
	Stationary (e.g., at home, office, school)	190	16.842	7.802	14.647	8.002

Two multiple regression models were estimated to examine whether patterns of primary LLM use predicted instrumental and relationship LLM dependency.

The model predicting instrumental LLM dependency accounted for a substantial proportion of variance ($R^2 = .41$, adjusted $R^2 = .40$). Frequency of LLM use showed strong and graded associations with instrumental dependency. Relative to the reference category (Less than 1-2 times per week), using the LLM 3-6 times per week was associated with higher instrumental dependency scores ($b = 4.15$, $SE = 0.76$, $t = 5.45$, $p < .001$, 95% CI [2.66, 5.65]), while multiple times daily use was associated with markedly higher instrumental dependency ($b = 10.20$, $SE = 0.80$, $t = 12.72$, $p < .001$, 95% CI [8.63, 11.78]).

Interaction modality was also a significant predictor of instrumental LLM dependency. Participants who interacted with the LLM using both text and voice reported higher instrumental dependency scores compared with text-only users ($b = 4.14$, $SE = 0.76$, $t = 5.48$, $p < .001$, 95% CI [2.65, 5.62]). In contrast, primary device type showed only a marginal association: using PC and other devices, relative to mobile devices, was associated with lower instrumental dependency, but this effect did not reach conventional levels of statistical significance ($b = -1.24$, $SE = 0.68$, $t = -1.82$, $p = .070$). Location of LLM use was not associated with instrumental dependency ($b = -0.02$, $SE = 0.66$, $t = -0.03$, $p = .972$).

The model predicting relationship LLM dependency explained a smaller proportion of variance ($R^2 = .30$, adjusted $R^2 = .29$). As in the instrumental model, frequency of LLM use emerged as a robust predictor. Using the LLM 3-6 times per week was associated with higher relationship dependency scores ($b = 2.88$, $SE = 0.82$, $t = 3.53$, $p < .001$, 95% CI [1.28, 4.49]), and multiple times daily use showed a large positive association ($b = 8.10$, $SE = 0.87$, $t = 9.36$, $p < .001$, 95% CI [6.40, 9.80]).

Unlike the instrumental model, primary device type was significantly associated with relationship dependency. Using PC and other devices, relative to mobile devices, was linked to lower relationship dependency scores ($b = -1.80$, $SE = 0.72$, $t = -2.48$, $p = .013$, 95% CI [-3.22, -0.37]). Interaction modality showed a similar pattern to the instrumental dimension: interacting with the LLM using both text and voice was associated with higher relationship dependency scores compared with text-only interaction ($b = 4.02$, $SE = 0.81$, $t = 4.96$, $p < .001$, 95% CI [2.43, 5.61]). Location of LLM use again showed no association with relationship dependency ($b = -0.25$, $SE = 0.73$, $t = -0.35$, $p = .727$).

Across both models, higher-frequency engagement with LLMs and multimodal interaction were consistently associated with stronger LLM dependency scores, whereas contextual factors such as location of use showed no association.

4. Discussion

The present study aimed to validate the German version of the Large Language Model Dependency Scale (LLM-D12-DE), a dual-dimensional instrument designed to assess both instrumental and relational forms of dependency on Large Language Models. The validation was conducted within a German-speaking sample characterized by active LLM engagement, ensuring that the construct of LLM

dependency was relevant and meaningful to the participants. The results provide robust evidence for the psychometric soundness of the LLM-D12-DE, replicating the original two-factor structure and demonstrating excellent internal consistency, convergent validity, and measurement invariance across gender.

Psychometric Properties of the LLM-D12-DE

The German version of the LLM-D12 demonstrated strong psychometric properties across multiple validation criteria, successfully replicating the factor structure observed in both the UK and Arabic validations (AlShakhsi et al., under review, Yankouskaya, Babiker, et al., 2025). This cross-cultural replication is particularly important given that psychometric structures can vary substantially across populations due to cultural differences in item interpretation, response styles, and construct manifestation (Borsboom, 2006; Van de Vijver & Tanzer, 2004). The consistency of factor structure across three linguistically and culturally distinct samples provides strong evidence for the universality of the instrumental-relational distinction in LLM dependency. Internal consistency estimates exceeded conventional thresholds for both subscales, with values consistent with those reported in prior validations (UK and Arab). The correspondence between classical reliability indices and model-based estimates indicates minimal multidimensionality within factors (Dunn et al., 2014; McNeish, 2018) and reinforces confidence in the unidimensional interpretation of each subscale. Average variance extracted values exceeded recommended thresholds, demonstrating that the latent constructs accounted for substantial proportions of variance in their respective indicators (Fornell & Larcker, 1981; Hair, 2009).

The reverse-coded item RD5 (*I use it solely as a tool, not to express my feelings or expect it to understand me*) consistently showed weaker factor loadings across all three validations, a pattern well-documented in the psychometric literature on reverse-coded items (Sonderer et al., 2013; Weijters & Baumgartner, 2012). Meta-analytic evidence demonstrates that reverse-coded items consistently show lower factor loadings and higher error variance due to increased cognitive load, careless responding, and the formation of distinct method factors (DiStefano & Motl, 2006; Huang et al., 2012; Podsakoff et al., 2003). Despite this weakness, the item's theoretical importance in capturing the absence of relational engagement, supports its retention. Future research might explore whether positively-worded items assessing low relational engagement perform more robustly while preserving construct coverage (Woods, 2006).

The establishment of scalar invariance across gender represents a crucial psychometric achievement, indicating that the LLM-D12-DE measures the same underlying constructs in male and female respondents and that observed group differences reflect true differences in latent dependency rather than measurement artifacts (Meredith, 1993; Vandenberg & Lance, 2000). Latent mean comparisons revealed that male participants reported higher instrumental dependency than females, aligning with prior research demonstrating gender differences in task-oriented technology engagement (Cai et al., 2017; Venkatesh & Morris, 2000).

Structural Relationships and Network Structure

Network analysis provided a complementary perspective on construct structure, revealing two densely interconnected clusters corresponding to the Instrumental and Relationship Dependency dimensions. This approach complements traditional latent variable modeling by examining pairwise conditional dependencies between observed variables and has been increasingly applied to understand symptom-level relationships in psychological assessment (Borsboom & Cramer, 2013; Christensen et al., 2024).

Within the Instrumental cluster, the items ID5 (*It is my go-to for assistance in decision-making*) and ID6 (*Making decisions without it feels somewhat uneasy*) emerged as the most central nodes, suggesting that habitual reliance on LLMs for decision support and discomfort when making independent decisions are core features of instrumental dependency. These findings align with theoretical models of cognitive offloading (Risko & Gilbert, 2016; Storm & Stone, 2015), which posit that individuals increasingly delegate cognitive tasks to external systems perceived as reliable and accessible. Empirical evidence demonstrates that while cognitive offloading reduces subjective mental effort and improves task performance under cognitive load (Dunn & Risko, 2016), repeated offloading may lead to reduced internal memory encoding, diminished problem-solving skills, and cognitive skill atrophy through disuse (Barr et al., 2015; Sparrow et al., 2011).

Within the Relationship cluster, RD3 (*It helps me feel less alone when I need to talk to someone*) emerged as the most central node overall, suggesting that loneliness reduction may be a primary psychological mechanism driving relational dependency on LLMs. This finding aligns with research on parasocial relationships (Horton & Richard Wohl, 1956; Tukachinsky & Stever, 2019) and anthropomorphism in human-computer interaction (Epley et al., 2007; Waytz et al., 2010), documenting that users can form one-sided relational bonds with conversational agents that mimic aspects of human companionship (Skjuve et al., 2021; Ta et al., 2020). Recent empirical work demonstrates that conversational AI can elicit self-disclosure comparable to human interaction, evoke emotional responses similar to those in human relationships, and provide perceived social support that buffers against loneliness (Ho et al., 2018; Liu & Sundar, 2018; Reis et al., 2017).

The relatively sparse cross-cluster connections indicate that these two forms of dependency, while strongly correlated at the latent level, are structurally distinct at the item level. This separation was further supported by the substantially worse fit of a single-factor model, with fit differences far exceeding conventional thresholds for meaningful model differences (Chen, 2007; Cheung & Rensvold, 2002). This structural arrangement suggests that a common underlying vulnerability toward LLM dependency may manifest through distinct pathways, with implications for both theory and intervention design. Users may become primarily instrumentally dependent, primarily relationally dependent, or show features of both, consistent with multifinality models in psychopathology (Cicchetti & Rogosch, 1996).

External Validation and Theoretical Integration

External validation analyses demonstrated that the LLM-D12-DE is sensitive to theoretically relevant psychological constructs, providing evidence for both convergent and discriminant validity. Both forms of dependency were consistently associated with greater internet addiction severity and more positive AI acceptance attitudes. The association with internet addiction suggests that LLM dependency may represent a specific manifestation of broader patterns of problematic digital engagement (Brand et al., 2016; Griffiths et al., 2016) rather than an entirely novel phenomenon, consistent with the component model of addiction (Griffiths, 2005). However, the unique association with perceived trustworthiness, combined with the distinct pattern of associations observed in the nomological network, indicates that LLM dependency is not reducible to general internet overuse and possesses distinct psychological antecedents, supporting its status as a discriminable construct (Strauss & Smith, 2009).

Perceived trustworthiness of LLMs predicted both forms of dependency, with stronger associations for instrumental dependency. This pattern is consistent with automation trust literature demonstrating that perceived reliability and predictability are primary determinants of trust in automated systems, and that trust is a prerequisite for delegation (Dzindolet et al., 2003; Hoff & Bashir, 2015; Lee & See, 2004; Merritt et al., 2013; Parasuraman & Riley, 1997). In the LLM context, trust is influenced by output accuracy, transparency of reasoning, consistency across interactions, and alignment with user values (Aroyo et al., 2021; Bussone et al., 2015; Jacovi et al., 2021). The weaker but significant association with relational dependency may reflect trust's role in fostering emotional openness and self-disclosure, foundational processes in parasocial bond formation (Bickmore et al., 2013; Følstad et al., 2018; Rubin & Step, 2000).

AI acceptance attitudes showed robust positive associations with both dependency forms, representing the largest effects observed. This finding is consistent with the Technology Acceptance Model and its extensions (Davis et al., 1989; Venkatesh & Davis, 2000; Venkatesh et al., 2003), with meta-analytic support across diverse technologies and cultural contexts (King & He, 2006; Schepers & Wetzels, 2007).

The null finding for AI fear attitudes warrants careful interpretation. It may reflect measurement limitations of the brief fear subscale (Sindermann et al., 2021), suggest that fear and acceptance represent orthogonal dimensions rather than opposite poles (Mitzner et al., 2019; Nomura et al., 2006). Alternatively, drawing on avoidance-based models of technology-related threat responses, it is conceivable that individuals with high AI-related fear refrain from using LLMs altogether and were therefore not captured by our inclusion criteria (Hunter et al., 1990; Liang & Xue, 2009; Sackett & Yang, 2000).

The finding that need for cognition predicts relational but not instrumental dependency supports the view that dependency on LLMs comprises distinct motivational dimensions rather than a single construct. While need for cognition is generally associated with engagement in effortful cognitive

activity, it may discourage relational dependency on LLMs, which offer limited opportunities for sustained cognitive elaboration compared to human interaction (Cacioppo & Petty, 1982; Nass, 1994). In contrast, instrumental dependency appears largely driven by situational and goal-related factors, such as task demands or efficiency considerations, rather than by intrinsic cognitive motivation (Kruglanski et al., 2002; Locke & Latham, 2002). These results highlight the importance of differentiating relational and instrumental forms of dependency in research on human-LLM interaction.

Usage Patterns as Dependency Predictors

The examination of LLM usage patterns as predictors of dependency contributes beyond psychological antecedents to examine how interactional characteristics shape dependency outcomes. Frequency of use emerged as the strongest predictor of both dependency forms, with a clear dose-response relationship. This finding is consistent with behavioral addiction models emphasizing habitual and frequent engagement in dependency development (Griffiths, 2005; LaRose et al., 2003), with extensive empirical support from research on internet, social media, and gaming addiction (Brand et al., 2014; Chou et al., 2005; Kuss & Griffiths, 2011; Lemmens et al., 2009; Müller et al., 2015). The graded increase across frequency levels suggests incremental dependency development through repeated exposure, consistent with habit formation theory (Lally et al., 2010; Verplanken & Aarts, 1999; Woods & Scott, 2016). Interaction modality showed significant associations with both dependency forms, with multimodal (text and voice) users reporting higher dependency than text-only users. This finding may reflect several mechanisms: multimodal interaction enhancing perceived naturalness and intimacy (Biocca et al., 2003; Daft & Lengel, 1986; Lee & See, 2004), voice input reducing cognitive effort and lowering barriers to frequent interaction (Ruan et al., 2018; Sweller, 1988), voice activating social schemas that prime perception of LLMs as human-like partners (Moon & Nass, 1996; Nass et al., 1994; Reeves & Nass, 1996), or multimodal users extracting more value through functional advantages (Davis et al., 1989). Experimental research manipulating modality while holding content constant would establish causal effects, while longitudinal studies could examine whether voice adoption accelerates dependency development. Device type showed differential associations: mobile users reported higher relational dependency than PC users, but device effects on instrumental dependency were only marginal. This pattern is theoretically meaningful as mobile devices are used in more intimate, personal contexts and offer constant availability (e Silva & Frith, 2012; Ling, 2012), potentially fostering emotional dependency (Cheever et al., 2014; Oulasvirta et al., 2012), whereas PC-based use may remain more task-focused (Haag et al., 1999; Hiltz, 2004). The physical and psychological intimacy of mobile devices may transfer to applications they host (Vincent, 2006; Weilenmann & Larsson, 2002). This interpretation aligns with Attachment Theory's emphasis on proximity and availability in emotional bond formation (Ainsworth, 1989; Mikulincer & Shaver, 2010). Location of use showed no association with either dependency form, contrasting with the UK validation. This discrepancy may reflect cultural differences in technology use norms, sampling differences, temporal factors, or genuine cross-cultural specificity.

Cross-Cultural Context and Mean Level Differences

Mean dependency scores in the German sample were notably lower than those in the Arabic validation and comparable to UK levels. This pattern aligns with cross-cultural research indicating that Germanic European populations exhibit more reserved AI attitudes compared to Middle Eastern populations (Gillespie et al., 2025; Liebherr et al., under review; Sindermann et al., 2021). Several cultural factors may contribute. Germany emphasizes data privacy, individual autonomy, and skepticism toward large-scale technological systems, institutionalized in GDPR and reflected in public discourse (Dinev et al., 2006; Floridi et al., 2018; Oyserman et al., 2002). Germany's high uncertainty avoidance (Hofstede, 2001) indicates cultural preference for rules, structure, and predictability, translating to greater caution in adopting unpredictable technologies (Doney et al., 1998; Yeniyurt & Townsend, 2003). This manifests in higher perceived risk, greater privacy concerns, and slower technology adoption compared to other nations (Dinev & Hart, 2006; Pavlou, 2003; Straub, 1994).

Germanic cultural values such as self-reliance, critical thinking, and skepticism may attenuate appeal of cognitive offloading or relational bonding with non-human agents. Germany's educational emphasis may foster resistance to cognitive outsourcing. In contrast, populations in collectivist regions with higher technology trust may be more receptive to integrating LLMs functionally and socially (Ameen & Willis, 2015; Hofstede, 1984; Srite & Karahanna, 2006). The cross-cultural variation underscores cultural context's importance in shaping human-AI interaction and cautions against assuming cross-cultural generalizability (Henrich et al., 2010; Matsumoto & Yoo, 2006). However, consistent psychometric properties across culturally distinct samples provide strong evidence for the universality of underlying constructs, even as prevalence varies (Norenzayan & Heine, 2005).

Implications for Theory, Practice, and Policy

The present validation has important theoretical and practical implications. Theoretically, the findings support the conceptual distinction between instrumental and relational LLM dependency, now replicated across three cultural contexts, suggesting this distinction reflects genuine psychological differences in how individuals relate to LLMs. This challenges purely instrumental technology use models by demonstrating relational factors constitute a distinct and meaningful engagement dimension with AI systems (Sundar & Lee, 2022). The dual-dimensional structure suggests interventions promoting healthy LLM use must address both cognitive and affective pathways. Educational initiatives focusing solely on critical evaluation may be insufficient without addressing emotional and relational dynamics underlying parasocial attachment (Skjuve et al., 2021; Zhai et al., 2024). In line with recent calls to integrate AI literacy into formal education curricula (Fußhöller et al., 2025), such initiatives should embed LLM literacy within broader digital well-being programs. These programs could aim to foster metacognitive awareness of cognitive offloading processes and develop affective regulation strategies to mitigate dependency risks. Comprehensive intervention programs might include cognitive components (critical evaluation, understanding AI limitations, maintaining cognitive skills), emotional

components (addressing loneliness, social skills development), and behavioral components (usage monitoring, structured reduction, alternative activities). Relative emphasis could be tailored to individual dependency profiles. For practitioners, the LLM-D12-DE provides a validated tool for assessing dependency in German-speaking populations, enabling targeted screening and monitoring. Educational institutions could implement routine screening to identify at-risk students, with elevated instrumental dependency triggering workshops on effective AI use and critical thinking, while high relational dependency might indicate need for social skills training or counseling addressing loneliness. Mental health professionals could incorporate the scale into intake assessments, particularly for clients presenting with potentially technology-related symptoms. Elevated relationship dependency might indicate LLM interactions substituting for human relationships, potentially exacerbating social isolation (Caplan, 2003; Morahan-Martin & Schumacher, 2003). For policymakers, findings highlight the need for evidence-based AI regulation accounting for psychological dimensions of human-AI interaction. While current EU debates emphasize transparency, accountability, and data protection (Floridi et al., 2018; Veale & Borgesius, 2021), regulatory frameworks should also consider potential for users to develop maladaptive dependencies. Policies might include mandatory dependency risk disclosures, design standards requiring healthy use features (usage dashboards, break reminders), educational campaigns promoting digital literacy, research funding monitoring dependency prevalence over time, and age-appropriate restrictions for minors. Policies should avoid paternalism, instead empowering informed decisions through proportionate governance (Koops et al., 2010). Importantly, dependency should not be equated with pathology absent evidence of negative impacts on well-being, functioning, or autonomy. The present study assessed dependency as a construct but not its consequences. High scorers might function well across life domains without distress, making autonomous technology choices. This aligns with debates about whether engagement frequency constitutes disorder or whether functional impairment must be demonstrated (Kardefelt-Winther, 2014; Van Rooij & Prause, 2014). Future research should examine whether dependency predicts adverse outcomes and identify clinically meaningful thresholds.

5. Limitations

Several limitations need to be acknowledged. First, the cross-sectional design precludes causal inference. While external validation demonstrated theoretically coherent associations, directionality remains unclear. The potential novelty effect (elevated dependency scores due to technology newness that may diminish with familiarity) requires longitudinal investigation. Second, reliance on self-report introduces possibilities of social desirability bias, recall bias, and common method variance. While CFA results provided no evidence of substantial common method effects, self-report correlates only moderately with objective usage measures. Future research should incorporate objective measures (e.g., logged interaction frequency, session duration, query types) to establish behavioral convergent validity. Multi-method approaches combining self-report, behavioral data, informant reports, and performance-based measures would provide robust construct validation. Third, online convenience sampling through

Prolific introduces self-selection bias and limits generalizability. Participants who voluntarily participate in LLM research may differ systematically in attitudes, usage patterns, and dependency levels. Prolific users represent a specific demographic. The sample's age and educational homogeneity may restrict variance on moderating variables. Future studies should focus on more diverse samples including older adults, individuals with lower digital literacy, rural populations, and clinical samples through mental health services. Fourth, LLM dependency remains in early developmental stages without consensus on diagnostic criteria, prevalence thresholds, or clinical significance. Unlike established constructs like internet gaming disorder or gambling disorder, LLM dependency lacks official recognition, validated cutoffs, or epidemiological data. Research is needed to determine whether elevated scores associate with functional impairment, distress, or psychopathology markers, and whether the construct meets behavioral addiction criteria.

6. Conclusion

The present study provides robust evidence for the validity and reliability of the German version of the Large Language Model Dependency Scale. The scale successfully replicates the dual-dimensional structure observed in prior validations, demonstrates excellent internal consistency and convergent validity, and shows measurement invariance across gender. Network analysis revealed structurally coherent clusters with interpretable centrality patterns. External validation confirmed theoretically coherent associations with relevant psychological constructs. Usage pattern analyses revealed systematic relationships between interaction characteristics and dependency levels, with frequency and multimodal interaction emerging as primary predictors. The availability of a psychometrically sound, culturally validated instrument for assessing LLM dependency in German-speaking populations represents an important step forward in the scientific study of human-AI interaction. As LLMs continue proliferating and becoming increasingly integrated into everyday life, the ability to systematically measure and monitor dependency will be essential for safeguarding user well-being, informing educational and clinical practice, and guiding evidence-based policy development. The LLM-D12-DE provides researchers, practitioners, and policymakers with a critical tool for understanding and addressing the psychological implications of this rapidly evolving technology.

Funding

The work was supported by Mercator Research Center Ruhr (Ko-2021-0023).

Data availability

The datasets generated by the survey research and codes for analysis are available in the Open Science Framework repository at <https://osf.io/j58mq/overview>. The author confirms that all data generated or analyzed during this study are included in this published article.

Conflict of interest

The authors declare no conflicts of interest.

Consent for publication

Not applicable.

Authors' contributions

ML: Conceptualized and designed the research, translated the original questionnaire and curated data, assisted with statistical analysis, and wrote the first draft. **AY:** Conceptualized and designed the research, designed, performed and reported statistical analysis, and wrote the first draft. **CM:** Included in the translation process, critically reviewed and edited the paper. **RA:** Conceptualized, designed and coordinated the research, translated the original questionnaire and curated data, acquired research funding, and critically reviewed and edited the paper.

Ethical Approval

This study was conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethical approval was obtained from Bournemouth University, UK (62239, 03/03/2025).

Informed Consent

Informed consent was obtained from all individual participants included in the study. Participants were informed about the purpose of the study, procedures, potential risks and benefits, and their right to withdraw at any time without penalty.

References

- Ainsworth, M. S. (1989). Attachments beyond infancy. *American psychologist*, 44(4), 709.
- AlShakhshi, S., Yankouskaya, A., Liebherr, M., Ali, R. (under review). Measuring Large Language Models Dependency: Validating the Arabic Version of the LLM-D12 Scale. *Arabian Journal for Science and Engineering*.
- Ameen, N., & Willis, R. (2015). The effect of cultural values on technology adoption in the Arab countries. *International Journal of Information Systems*, 2, 1-7.
- Aroyo, A. M., De Bruyne, J., Dheu, O., Fosch-Villaronga, E., Gudkov, A., Hoch, H., Jones, S., Lutz, C., Sætra, H., & Solberg, M. (2021). Overtrusting robots: Setting a research agenda to mitigate overtrust in automation. *Paladyn, Journal of Behavioral Robotics*, 12(1), 423-436.
- Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior*, 48, 473-480.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588.
- Bickmore, T. W., Vardoulakis, L. M. P., & Schulman, D. (2013). Tinker: a relational agent museum guide. *Autonomous agents and multi-agent systems*, 27, 254-276.
- Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & Virtual Environments*, 12(5), 456-480.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425-440.
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, 9(1), 91-121.
- Brand, M., Laier, C., & Young, K. S. (2014). Internet addiction: coping styles, expectancies, and treatment implications. *Frontiers in psychology*, 5, 1256.
- Brand, M., Young, K. S., Laier, C., Wölfling, K., & Potenza, M. N. (2016). Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An Interaction of Person-Affect-Cognition-Execution (I-PACE) model. *Neuroscience & Biobehavioral Reviews*, 71, 252-266.
- Brauner, P., Glawe, F., Liehner, G. L., Vervier, L., & Ziefle, M. (2024). AI perceptions across cultures: Similarities and differences in expectations, risks, benefits, tradeoffs, and value in Germany and China. *arXiv preprint arXiv:2412.13841*.
- Bussone, A., Stumpf, S., & O'Sullivan, D. (2015). *The role of explanations on trust and reliance in clinical decision support systems*. In 2015 International Conference on Healthcare Informatics (ICHI) (pp. 160-169)
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, 42(1), 116.
- Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, 105, 1-13.
- Caplan, S. E. (2003). Preference for online social interaction: A theory of problematic Internet use and psychosocial well-being. *Communication research*, 30(6), 625-648.
- Cheever, N. A., Rosen, L. D., Carrier, L. M., & Chavez, A. (2014). Out of sight is not out of mind: The impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users. *Computers in Human Behavior*, 37, 290-297.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Chou, C., Condrón, L., & Belland, J. C. (2005). A review of the research on Internet addiction. *Educational psychology review*, 17(4), 363-388.
- Christensen, A. P., Garrido, L. E., Guerra-Peña, K., & Golino, H. (2024). Comparing community detection algorithms in psychometric networks: A Monte Carlo simulation. *Behavior Research Methods*, 56(3), 1485-1505.
- Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and psychopathology*, 8(4), 597-600.

- Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, 32(5), 554-571.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, 35(8), 982-1003.
- Dergaa, I., Ben Saad, H., Glenn, J. M., Amamou, B., Ben Aissa, M., Guelmami, N., Fekih-Romdhane, F., & Chamari, K. (2024). From tools to threats: a reflection on the impact of artificial-intelligence chatbots on cognitive health. *Frontiers in psychology*, 15, 1259845.
- Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., & Colautti, C. (2006). Privacy calculus model in e-commerce—a study of Italy and the United States. *European journal of information systems*, 15(4), 389-402.
- Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61-80.
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13(3), 440-464.
- Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. *Academy of management review*, 23(3), 601-620.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, 105(3), 399-412.
- Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive science*, 40(5), 1080-1127.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koochang, A., Raghavan, V., & Ahuja, M. (2023). Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697-718.
- e Silva, A. d. S., & Frith, J. (2012). *Mobile interfaces in public spaces: Locational privacy, control, and urban sociability*. Routledge.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4), 864.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, 28(4), 689-707.
- Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). *What makes users trust a chatbot for customer service? An exploratory interview study*. In Proceedings of the 6th International Conference on Internet Science (INSCI 2018) (pp. 194–208).
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 18(1), 39-50.
- Fußhöller, A., Lechner, F., Schlicker, N., Muehlensiepen, F., Mayr, A., Kuhn, S., Hirsch, M. C., & Knitza, J. (2025). Perceptions, Usage, and Educational Impact of ChatGPT Among Medical Students in Germany: Cross-Sectional Mixed Methods Survey. *JMIR Formative Research*, 9(1), e81484.
- Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). *Trust, attitudes and use of artificial intelligence*. The University of Melbourne & KPMG.
- Giubilini, A., Porsdam Mann, S., Voinea, C., Earp, B., & Savulescu, J. (2024). Know thyself, improve thyself: personalized LLMs for self-knowledge and moral enhancement. *Science and Engineering Ethics*, 30(6), 54.
- Griffiths, M. (2005). A ‘components’ model of addiction within a biopsychosocial framework. *Journal of Substance use*, 10(4), 191-197.
- Griffiths, M. D., Kuss, D. J., Billieux, J., & Pontes, H. M. (2016). The evolution of Internet addiction: A global perspective. *Addictive behaviors*, 53, 193-195.

- Haag, S., Cummings, M., & Dawkins, J. (1999). *Management information systems for the information age*. McGraw-Hill Higher Education.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3), 61-83.
- Hiltz, S.-R. (2004). Predicting learning from asynchronous online discussions. *Journal of asynchronous learning networks*.
- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of communication*, 68(4), 712-733.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407-434.
- Hofstede, G. (1984). *Culture's consequences: International differences in work-related values* (Vol. 5). Sage publications.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.
- Horton, D., & Richard Wohl, R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, 19(3), 215-229.
- Hu, B., Mao, Y., & Kim, K. J. (2023). How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. *Computers in Human Behavior*, 145, 107760.
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114.
- Huang, Y., Sun, L., Wang, H., Wu, S., Zhang, Q., Li, Y., Gao, C., Huang, Y., Lyu, W., & Zhang, Y. (2024). Trustllm: Trustworthiness in large language models. *arXiv preprint arXiv:2401.05561*.
- Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75(1), 28.
- Ibeh, L., Mutai, N. C., Popoola, O. M., Cuong, N. M., & Ejiofor, S. (2025). Exploring perspectives on ChatGPT integration in education: A student-centered study of benefits, concerns, and global implications for responsible AI integration. *Research in Learning Technology*, 33.
- Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). *Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI*. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21) (pp. 624–635). ACM.
- Kardefelt-Winther, D. (2014). A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. *Computers in Human Behavior*, 31, 351-354. <https://doi.org/https://doi.org/10.1016/j.chb.2013.10.059>
- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & management*, 43(6), 740-755.
- Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, 18(1), rm1.
- Koops, B.-J., Hildebrandt, M., & Jaquet-Chiffelle, D.-O. (2010). Bridging the accountability gap: Rights for new entities in the information society? *Minn. JL Sci. & Tech.*, 11, 497.
- Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Woo Young, C., & Sleeth-Keppler, D. (2002). A theory of goal systems. In *Advances in Experimental Social Psychology* (Vol. 34, pp. 331-378). Academic Press.
- Kuss, D. J., & Griffiths, M. D. (2011). Online social networking and addiction—a review of the psychological literature. *International journal of environmental research and public health*, 8(9), 3528-3552.
- Kuss, D. J., Griffiths, M. D., & Binder, J. F. (2013). Internet addiction in students: Prevalence and risk factors. *Computers in Human Behavior*, 29(3), 959-966.
- Lally, P., Van Jaarsveld, C. H., Potts, H. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European journal of social psychology*, 40(6), 998-1009.
- LaRose, R., Lin, C. A., & Eastin, M. S. (2003). Unregulated Internet usage: Addiction, habit, or deficient self-regulation? *Media Psychology*, 5(3), 225-253.

- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50-80.
- Lemmens, J. S., Valkenburg, P. M., & Peter, J. (2009). Development and validation of a game addiction scale for adolescents. *Media Psychology*, 12(1), 77-95.
- Liang, H., & Xue, Y. (2009). Avoidance of information technology threats: A theoretical perspective. *MIS quarterly*, 71-90.
- Liebherr, M., Almourad, M.B., Yankouskaya, A., Thomas, J., Guandong, X., & Ali, R. (under review). Psychological Needs and AI Delegation Across Four Social Domains - A Cross-Cultural Analysis of 35 Nations. *Telematics and Informatics*.
- Ling, R. (2012). *Taken for grantedness: The embedding of mobile communication into society*. MIT press.
- Lins de Holanda Coelho, G., HP Hanel, P., & J. Wolf, L. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8), 1870-1885.
- Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, 21(10), 625-636.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, 57(9), 705.
- Matsumoto, D., & Yoo, S. H. (2006). Toward a new generation of cross-cultural research. *Perspectives on Psychological Science*, 1(3), 234-250.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological methods*, 23(3), 412.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human factors*, 55(3), 520-534.
- Mikulincer, M., & Shaver, P. R. (2010). *Attachment in adulthood: Structure, dynamics, and change*. Guilford Publications.
- Mitzner, T. L., Savla, J., Boot, W. R., Sharit, J., Charness, N., Czaja, S. J., & Rogers, W. A. (2019). Technology adoption by older adults: Findings from the PRISM trial. *The Gerontologist*, 59(1), 34-44.
- Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication research*, 23(6), 651-674.
- Morahan-Martin, J., & Schumacher, P. (2003). Loneliness and social uses of the Internet. *Computers in Human Behavior*, 19(6), 659-671.
- Müller, K. W., Janikian, M., Dreier, M., Wölfling, K., Beutel, M. E., Tzavara, C., Richardson, C., & Tsitsika, A. (2015). Regular gaming behavior and internet gaming disorder in European adolescents: results from a cross-national representative survey of prevalence, predictors, and psychopathological correlates. *European child & adolescent psychiatry*, 24(5), 565-574.
- Nass, C. (1994). Knowledge or skills: which do administrators learn from experience? *Organization Science*, 5(1), 38-50.
- Nass, C., Steuer, J., & Tauber, E. R. (1994). *Computers are social actors*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94) (pp. 72-78). ACM
- Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). *Measurement of anxiety toward robots*. ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication,
- Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological bulletin*, 131(5), 763.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.
- Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous computing*, 16(1), 105-114.
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological bulletin*, 128(1), 3.

- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2), 230-253.
- Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, 7(3), 101-134.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879.
- Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, 10(10), 19-36.
- Reis, A., Paulino, D., Paredes, H., & Barroso, J. (2017). *Using intelligent personal assistants to strengthen the elderly's social bonds: A preliminary evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri*. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction (UAHCI 2017). Springer.
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences*, 20(9), 676-688.
- Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. A. (2018). Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4), 1-23.
- Rubin, A. M., & Step, M. M. (2000). Impact of motivation, attraction, and parasocial interaction on talk radio listening. *Journal of Broadcasting & Electronic Media*, 44(4), 635-654.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: an expanded typology. *Journal of Applied Psychology*, 85(1), 112.
- Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & management*, 44(1), 90-103.
- Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI-Künstliche Intelligenz*, 35, 109-118.
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149, 102601.
- Sonderen, E. v., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS one*, 8(7), e68967.
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776-778.
- Srite, M., & Karahanna, E. (2006). The role of espoused national cultural values in technology acceptance. *MIS quarterly*, 679-704.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and individual differences*, 42(5), 893-898.
- Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological science*, 26(2), 182-188.
- Straub, D. W. (1994). The Effect of Culture on IT Diffusion: E-Mail and FAX in Japan and the US. *Information Systems Research*, 5(1), 23-47.
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, 5, 1-25.
- Sundar, S. S., & Lee, E.-J. (2022). Rethinking communication in the era of artificial intelligence. *Human Communication Research*, 48(3), 379-385.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.
- Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of medical Internet research*, 22(3), e16235.
- Tukachinsky, R., & Stever, G. (2019). Theorizing development of parasocial engagement. *Communication Theory*, 29(3), 297-318.

- Valenti, G. D., Craparo, G., & Faraci, P. (2025). The development of a short version of the internet addiction test: the IAT-7. *International Journal of Mental Health and Addiction*, 23(2), 1028-1053.
- Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European review of applied psychology*, 54(2), 119-135.
- Van Rooij, A., & Prause, N. (2014). A critical review of "Internet addiction" criteria with suggestions for the future. *Journal of Behavioral Addictions*, 3(4), 203-213.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70.
- Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU artificial intelligence act. *arXiv preprint arXiv:2107.03721*.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS quarterly*, 115-139.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- Verplanken, B., & Aarts, H. (1999). Habit, attitude, and planned behaviour: is habit an empty construct or an interesting case of goal-directed automaticity? *European review of social psychology*, 10(1), 101-134.
- Vincent, J. (2006). Emotional attachment and mobile phones. *Knowledge, Technology & Policy*, 19(1), 39-44.
- Wartberg, L., Kriston, L., & Thomasius, R. (2020). Internet gaming disorder and problematic social media use in a representative sample of German adolescents: Prevalence estimates, comorbid depressive symptoms and related psychosocial aspects. *Computers in Human Behavior*, 103, 31-36.
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219-232.
- Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of marketing research*, 49(5), 737-747.
- Weilenmann, A., & Larsson, C. (2002). Local use and sharing of mobile phones. In *Wireless world: Social and interactional aspects of the mobile age* (pp. 92-107). Springer.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186-191.
- Woods, H. C., & Scott, H. (2016). # Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of adolescence*, 51, 41-49.
- Yankouskaya, A., Babiker, A. B., Rizvi, S. W., Alshakhsi, S., Liebherr, M., & Ali, R. (2025). LLM-D12: A Dual-Dimensional Scale of Instrumental and Relational Dependencies on Large Language Models. *arXiv preprint arXiv:2506.06874*.
- Yankouskaya, A., Liebherr, M., & Ali, R. (2025). Can ChatGPT Be Addictive? A Call to Examine the Shift from Support to Dependence in AI Conversational Large Language Models. *Human-Centric Intelligent Systems*, 1-13. <https://doi.org/10.1007/s44230-025-00090-w>
- Yeniyurt, S., & Townsend, J. D. (2003). Does culture explain acceptance of new products in a country? An empirical investigation. *International Marketing Review*, 20(4), 377-396.
- Yu, S.-C., Chen, H.-R., & Yang, Y.-W. (2024). Development and validation the Problematic ChatGPT Use Scale: a preliminary report. *Current Psychology*, 43(31), 26080-26092.
- 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.
- Zhang, X., Yin, M., Zhang, M., Li, Z., & Li, H. (2025). The Development and Validation of an Artificial Intelligence Chatbot Dependence Scale. *Cyberpsychology, Behavior, and Social Networking*, 28(2), 126-131.

- Zhang, Z., Rossi, R. A., Kveton, B., Shao, Y., Yang, D., Zamani, H., Derroncourt, F., Barrow, J., Yu, T., & Kim, S. (2024). Personalization of large language models: A survey. *arXiv preprint arXiv:2411.00027*.
- Zhu, M., Weng, Y., Yang, L., & Zhang, Y. (2024). Personality alignment of large language models. *arXiv preprint arXiv:2408.11779*.

Appendix 1

Large Language Model Dependency Questionnaire (LLM-D12) - The English Version

Thinking specifically about the large language model (LLM) you use most often (such as ChatGPT, DeepSeek, Gemini, Copilot, etc.), please indicate the extent to which you agree or disagree with each of the following statements:

Item nr	item	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
ID1	Without it, I feel less confident when making decisions						
ID2	I use it sometimes without realizing how much time I spend immersed in it						
ID3	I feel much more rewarded and pleased when completing tasks using it						
ID4	I turn to it for support in decisions, even when I can make them myself with some effort						
ID5	It is my go-to for assistance in decision-making						
ID6	Making decisions without it feels somewhat uneasy						
RD1	I share details about my private life with it						
RD2	I interact with it as if it were a genuine companion						
RD3	It helps me feel less alone when I need to talk to someone						
RD4	It adds to my social life, making socializing more engaging and interesting						
RD5 [R]	I use it solely as a tool, not to express my feelings or expect it to understand me						

RD6	It helps me feel less alone, reducing the need to talk to others						
-----	--	--	--	--	--	--	--

Scoring:

Items are rated on a 6-point scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, 6 = Strongly Agree.

[R] indicates that the item shall be reversed when scoring

ID – Instrumental Dependency, RD – Relationship Dependency

Instrumental and Relationship Dependency should each be calculated as the sum of responses on their respective items. The two scales and their scores can be used separately. It is also possible to use the total score for LLM Dependency, obtained by summing the scores of both scales.

Notes:

The LLM examples in the introductory context can be updated to reflect either the most popular models at the time or those most relevant to the context in which the scale is administered.

The scale may be used at the item level, subject to several considerations and limitations. Please refer to Appendix 2 of the English version of the scale, accessible at: <https://doi.org/10.48550/arXiv.2506.06874>

Large Language Model Dependency Questionnaire (LLM-D12-DE) - The German Version

Denken Sie an das Sprachmodell (LLM), das Sie am häufigsten verwenden (z. B. ChatGPT, DeepSeek, Gemini, Copilot usw.), Bitte geben Sie an, in welchem Ausmaß Sie den folgenden Aussagen zustimmen bzw. sie ablehnen.

Item Nr	Item	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme eher nicht zu	Stimme eher zu	Stimme zu	Stimme voll und ganz zu
ID1	Wenn ich Entscheidungen ohne mein primäres LLM treffe, fühle ich mich unsicher.						
ID2	Ich benutze mein primäres LLM manchmal, ohne zu merken, wie viel Zeit ich damit verbringe.						
ID3	Es macht mir deutlich mehr Freude und gibt mir ein besseres Gefühl, wenn ich Aufgaben mit meinem primären LLM erledige.						

ID4	Ich nutze mein primäres LLM zur Unterstützung bei Entscheidungen, die ich mit etwas Mühe auch selbst treffen könnte.						
ID5	Mein primäres LLM ist meine erste Anlaufstelle, wenn ich Hilfe bei der Entscheidungsfindung brauche.						
ID6	Ohne mein primäres LLM Entscheidungen zu treffen, verunsichert mich etwas.						
RD1	Ich teile Details aus meinem Privatleben mit meinem primären LLM.						
RD2	Ich interagiere mit meinem primären LLM so, als wäre es eine reale Bezugsperson.						
RD3	Mein primäres LLM hilft mir, mich weniger allein zu fühlen, wenn ich mit jemandem reden muss.						
RD4	Mein primäres LLM bereichert mein soziales Leben und macht soziale Kontakte ansprechender und interessanter.						
RD5 [R]	Ich benutze mein primäres LLM nur als Hilfsmittel, nicht um meine Gefühle auszudrücken oder zu erwarten, dass es mich versteht.						
RD6	Mein primäres LLM hilft mir, mich weniger allein zu fühlen, und verringert das Bedürfnis, mit anderen zu sprechen.						

Auswertung:

Die Items werden auf einer 6-stufigen Skala bewertet: 1 = Stimme überhaupt nicht zu, 2 = Stimme nicht zu, 3 = Stimme eher nicht zu, 4 = Stimme eher zu, 5 = Stimme zu, 6 = Stimme voll und ganz zu.

[R] kennzeichnet Items, die bei der Auswertung invertiert (revers kodiert) werden müssen.

ID = Instrumentelle Abhängigkeit, RD = Beziehungsabhängigkeit.

Instrumentelle Abhängigkeit und Beziehungsabhängigkeit sollten jeweils als Summe der Antworten auf die zugehörigen Items berechnet werden. Beide Skalen sowie ihre jeweiligen Werte können getrennt verwendet werden. Alternativ kann ein Gesamtwert für die LLM-Abhängigkeit gebildet werden, indem die Werte beider Skalen aufsummiert werden.

Anmerkungen:

Die in der Einleitung genannten Beispiele für LLMs können aktualisiert werden, um entweder die jeweils populärsten Modelle oder jene widerzuspiegeln, die für den jeweiligen Anwendungskontext der Skala am relevantesten sind.

Die Skala kann auch auf Itemebene verwendet werden, wobei mehrere Aspekte und Einschränkungen zu berücksichtigen sind.

Bitte beachten Sie hierzu Anhang 2 der englischen Version der Skala, abrufbar unter: <https://doi.org/10.48550/arXiv.2506.06874>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix1.docx](#)