1/2026, Volume 18, Issue 1
ISSN 1803-537X

BARRIERS OF ARTIFICIAL INTELLIGENCE IN SOLVING LOGIC PROBLEMS FOR YOUNGER SCHOOL-AGE STUDENTS

Martina UHLÍŘOVÁ*, Univerzita Palackého v Olomouci, Česká republika
Jitka LAITOCHOVÁ, Univerzita Palackého v Olomouci, Česká republika
Lucie ŠTAFOVÁ, Univerzita Palackého v Olomouci, Česká republika
Veronika VESELÁ, Univerzita Palackého v Olomouci, Česká republika

Přijato: 9. 2. 2026 / Akceptováno: 17. 8. 2026
Typ článku: Výzkumná studie
DOI: 10.5507/jtie.2026.009

Abstract: This paper focuses on the analysis of errors in mathematical problem-solving generated by selected artificial intelligence systems, with particular attention to tasks designed for younger primary school pupils. The study is based on a set of 144 problems from the Mathematical Kangaroo competition, with a detailed analysis of the Ecolier category. The tasks were submitted to three generative AI systems (Gemini 2.0 Flash, ChatGPT, and Microsoft 365 Copilot). The study aimed to compare the performance of individual AI systems and to identify the structure and causes of errors in their solutions. The results reveal statistically significant differences between the systems and highlight increased error rates in geometric and context-based logical tasks. These findings are interpreted through the lens of embodied cognition, which helps to explain the discrepancy between formally correct but contextually inadequate AI solutions and the intuitive reasoning strategies employed by children.

Key words: artificial intelligence, Mathematical Kangaroo, Ecolier category, word problems, primary education.

BARIÉRY UMĚLÉ INTELIGENCE PŘI ŘEŠENÍ LOGICKÝCH ÚLOH PRO ŽÁKY MLADŠÍHO ŠKOLNÍHO VĚKU

Abstrakt: Článek se zabývá analýzou chyb v řešeních matematických úloh generovaných vybranými systémy umělé inteligence se zaměřením na úlohy určené žákům mladšího školního věku. Výzkum vychází ze souboru 144 úloh soutěže Matematický

*Autor pro korespondenci: martina.uhlirova@upol.cz

klokan, přičemž podrobná analýza je věnována úlohám kategorie Klokánek. Úlohy byly zadány třem systémům generativní umělé inteligence (Gemini 2.0 Flash, ChatGPT a Microsoft 365 Copilot) v tzv. zero-shot režimu. Cílem studie bylo porovnat výkon jednotlivých AI systémů a identifikovat strukturu a příčiny chyb v jejich řešeních. Výsledky ukazují statisticky významné rozdíly mezi jednotlivými systémy a zároveň poukazují na zvýšenou chybovost v úlohách s geometrickým a logickým kontextem. Tyto chyby jsou interpretovány v rámci konceptu vtěleného poznání, který umožňuje vysvětlit rozdíly mezi formálně správným, avšak kontextově neadekvátním řešením AI a intuitivním řešením dětí.

Klíčová slova: umělá inteligence, Matematický klokan, kategorie Klokánek, slovní úlohy, 1. stupeň ZŠ.

1 Úvod

Rychlý rozvoj nástrojů založených na umělé inteligenci (AI) v posledních letech výrazně proměňuje vzdělávací prostředí a otevírá nové otázky týkající se výuky i učení matematiky. Problematice využití umělé inteligence ve vzdělávání, včetně výuky dětí mladšího školního věku, se v posledních letech věnuje řada studií (např. UNESCO, 2023), které upozorňují nejen na potenciál těchto nástrojů, ale i na rizika spojená s nekritickým přebíráním generovaných řešení. Generativní nástroje AI jsou dnes schopny řešit matematické úlohy, vysvětlovat postupy a poskytovat okamžitou zpětnou vazbu, což vyvolává zájem o jejich možné využití již na 1. stupni základní školy. Zároveň však vyvstávají otázky týkající se limitů těchto nástrojů, zejména v oblastech, kde matematické řešení úloh úzce souvisí s porozuměním kontextu, vizuální interpretací a intuitivním uvažováním.

V posledních letech vznikla řada studií zaměřených na schopnost velkých jazykových modelů (Large Language Models, LLM) řešit matematické úlohy a provádět logické usuzování. Výsledky těchto studií ukazují, že současné generativní modely dosahují vysoké úspěšnosti v úlohách založených na standardních algoritmických postupech, avšak jejich výkon výrazně klesá v situacích vyžadujících práci s implicitními informacemi, vizuální reprezentací nebo flexibilní změnou strategie řešení (Hendrycks et al., 2021; Wei et al., 2022). Výzkumy rovněž naznačují, že úspěšné řešení matematické úlohy nemusí být podmíněno pouze správným výpočtem, ale také schopností interpretovat kontext zadání a identifikovat relevantní vztahy mezi jednotlivými prvky problému (Kambhampati, 2024). Další studie upozorňují na přetrvávající obtíže při řešení úloh vyžadujících práci

1/2026, Volume 18, Issue 1
ISSN 1803-537X

s reálným kontextem, prostorovou představivostí nebo vícekrokovou dedukcí (Strohmaier et al., 2025; Boye & Moell, 2025).

Specifickou oblast představují úlohy založené na vizuálním a prostorovém uvažování. Přestože moderní multimodální modely dokáží zpracovávat obrazová data, řada studií upozorňuje na přetrvávající obtíže při interpretaci prostorových vztahů, mentální manipulaci s objekty nebo identifikaci implicitních pravidel grafického zadání (Xu et al., 2025). Tyto poznatky naznačují, že některé typy úloh běžně řešené žáky mladšího školního věku mohou představovat pro generativní AI specifickou výzvu.

Zvláštní pozornost si v tomto ohledu zaslouží matematické slovní a kontextové úlohy, které jsou považovány za klíčový prostředek rozvoje matematické gramotnosti. Tyto úlohy nevyžadují pouze aplikaci formálních početních postupů a algoritmů, ale také schopnost porozumět zadání, identifikovat relevantní informace, vytvářet vhodné reprezentace a reflektovat smysluplnost výsledku (Schoenfeld, 2016). Právě v této kombinaci jazykových, logických a vizuálních nároků se mohou projevit rozdíly mezi strategií řešení člověka a řešením generovaným umělou inteligencí.

V současnosti je pozornost oborových didaktiků směřována k možnostem a limitům využití AI ve vzdělávání dětí mladšího školního věku. Výzkumné práce upozorňují na potenciál personalizace učení a podpory učitele, ale zároveň zdůrazňují potřebu kritického a didakticky ukotveného využívání AI nástrojů (Bártek et al., 2025). V oblasti matematického vzdělávání na 1. stupni ZŠ však zatím chybí detailnější analýza příčiny selhání AI při řešení matematických úloh a identifikace těchto chybových vzorců, která může přispět k hlubšímu porozumění rozdílům mezi lidským a strojovým řešením matematických problémů.

Cílem tohoto článku je analyzovat chybovost vybraných systémů generativní umělé inteligence při řešení matematických úloh a zaměřit se na strukturu a příčiny těchto chyb. Studie vychází z úloh mezinárodní soutěže Matematický klokan, konkrétně z kategorie Klokánek určené žákům 4. a 5. ročníku základní školy. Pozornost je věnována zejména úlohám s geometrickým a logickým kontextem, u nichž se opakovaně ukazuje význam vizuální interpretace a heuristických strategií.

1/2026, Volume 18, Issue 1
ISSN 1803-537X

2 Metodologie výzkumu

Výzkum byl zaměřen na analýzu řešení matematických úloh generovaných vybranými systémy umělé inteligence. Hlavním cílem bylo porovnat výkon jednotlivých AI systémů mezi sebou a identifikovat strukturu chyb v jejich řešeních. Důraz byl kladen na kvalitativní analýzu chyb a na hledání jejich možných interpretačních příčin, nikoli na srovnání výkonu AI s výkonem žáků. Naše pozornost se soustředila na otázku, v jakých typech úloh se chyby AI objevují nejčastěji a jakého jsou charakteru. S ohledem na obecnou dostupnost byly zvoleny následující tři systémy generativní umělé inteligence:

- Google Gemini – model 2.0 Flash (AI1),
- Open AI ChatGPT (AI2),
- Microsoft 365 Copilot (AI3).

Do výzkumného šetření bylo celkem 144 úloh (S0). Soubor úloh zahrnoval úlohy z české verze soutěže Matematický klokan z let 2023 a 2024. Byly vybrány všechny úlohy kategorií Klokánek, Benjamín a Kadet. Úlohy byly z oblasti aritmetiky, algebry, geometrie a logiky. Pro podrobnější analýzu v této studii byly zvoleny soutěžní úlohy kategorie Klokánek (n = 24) z roku 2024 (S1), které jsou určeny žákům 4. a 5. ročníku základní školy. (Zastoupení jednotlivých typů úloh je zachyceno v Tabulce 2.) Úlohy této kategorie byly zvoleny záměrně, neboť kombinují relativně jednoduché početní operace s vyššími nároky na porozumění zadání, vizuální interpretaci a práci s implicitními pravidly. Právě tyto charakteristiky činí úlohy kategorie Klokánek vhodným materiálem pro analýzu limitů generativní umělé inteligence.

3 Realizace výzkumného šetření

Úlohy byly AI systémům zadány v tzv. zero-shot promptingu, tedy bez uvedení vzorového řešení, nápovědy či dodatečných instrukcí, pouze s pokynem „Vyřeš úlohu“. Tento postup odpovídá běžnému uživatelskému využití AI nástrojů a zároveň reflektuje metodologický rámec (Bártek et al., 2025).

V případě, že AI systém poskytl nesprávné řešení, následoval první doplňující prompt: „Zkontroluj své řešení ještě jednou. Neobsahuje některý krok chybu?“ Pokud AI na chybném řešení setrvávala, byl použit druhý doplňující prompt: „Zkus při řešení využít jinou strategii.“ Cílem bylo upřesnění zadání nebo upozornění

na nesoulad výsledku, nikoli přímé navedení ke správnému řešení. Tento postup umožnil sledovat, zda AI dokáže revidovat původní strategii, nebo zda u ní přetrvává chybné pojetí úlohy.

Získaná řešení byla analyzována z hlediska správnosti a dále kvalitativně kategorizována podle typu chyby (např. chybná interpretace zadání, nevhodná volba strategie, ignorování vizuálního kontextu). U vybraných úloh byla následně realizovaná další komunikace s AI, která umožnila detailnější vhled do mechanismů vzniku chyby.

4 Zjištěné výsledky a jejich interpretace

Analýza dat ukázala, že v kompletním souboru úloh (n = 144) i v kategorii Klokánek (n = 24) dosahovaly jednotlivé AI systémy rozdílné úspěšnosti. V obou případech vykazoval největší úspěšnost řešení Open AI ChatGPT (89,58 % pro S0 a 95,83 % pro S1). Největší rozdíl v úspěšnosti řešení identifikujeme u Microsoft 365 Copilot (AI3). Výsledky jsou shrnuty v tabulce 1.

	n	AI1 [%]	AI2 [%]	AI3 [%]
Kompletní soubor (S0)	144	75,00	**89,58**	75,69
Klokánek (S1)	24	79,17	**95,83**	66,67
Δ		4,17	**6,25**	9,02

Legenda: n – počet úloh v souboru, Δ – absolutní rozdíl

Tabulka 1 Úspěšnost řešení AI1, AI2, AI3

Rozdíly v úspěšnosti řešení mezi jednotlivými AI systémy byly testovány pomocí χ^2 testu nezávislosti. V kategorii Klokánek (n = 24) byly zjištěny statisticky významné rozdíly mezi systémy AI1, AI2 a AI3 ($\chi^2 = 6{,}56$; $p = 0{,}038$). Rovněž v celém souboru úloh (n = 144) se rozdíly v úspěšnosti ukázaly jako statisticky významné, a to na vyšší hladině významnosti ($\chi^2 = 14{,}33$; $p < 0{,}001$). Tyto výsledky potvrzují, že typ použitého AI systému má významný vliv na úspěšnost řešení matematických úloh daného typu.

V tomto článku byla provedena analýza chybovosti samostatně pro úlohy kategorie Klokánek (n = 24). Cílem analýzy bylo porovnání výkonu jednotlivých AI systémů mezi sebou a identifikace struktury chyb v řešeních generovaných

umělou inteligencí, se zvláštním důrazem na úlohy s vyšší mírou kontextové, vizuální a logické náročnosti.

	n	AI1 [%]	AI2 [%]	AI3 [%]
Kompletní soubor (S0)	144	75,00	**89,58**	75,69
Klokánek (S1)	24	79,17	**95,83**	66,67
Δ		4,17	**6,25**	9,02

Legenda: n – počet úloh v souboru, Δ – absolutní rozdíl

Tabulka 1 Úspěšnost řešení AI1, AI2, AI3

Soubor úloh	n	AI1 [%]	AI2 [%]	AI3 [%]
Klokánek (S1)	24	20,83	4,17	33,33
Geometrické úlohy	6	50,00	0,00	16,67
Algebraické úlohy	2	50,00	0,00	0,00
Aritmetické úlohy	7	0,00	0,00	28,57
Logické úlohy	9	11,11	11,11	55,56

Legenda: n – počet úloh v souboru

Tabulka 2 Chybovost řešení AI1, AI2, AI3

Z hlediska struktury chyb se ukázalo, že chyby AI systémů se v kategorii Klokánek výrazně soustředily do úloh s geometrickým a logickým kontextem. Tyto úlohy vyžadovaly nejen korektní aritmetické operace, ale především porozumění vizuálnímu uspořádání, implicitním pravidlům zadání a schopnost flexibilně přizpůsobit řešitelskou strategii specifikům konkrétní situace. Právě v těchto aspektech se ukázala omezení současných generativních modelů.

Podrobnější analýza vybraných úloh kategorie Klokánek ukázala, že chybné řešení AI často nevzniká v důsledku izolovaného numerického pochybení, ale jako důsledek nesprávně zvolené interpretační strategie, která přetrvává i při opakované interakci. Tento jev byl detailně sledován u vybraných úloh, u nichž byla vedena další komunikace s AI systémem.

Přestože AI disponuje rozsáhlým aparátem pro výpočetní operace, při řešení některých kontextových úloh naráží vybrané modely generativní AI na specifické

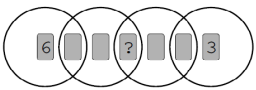
1/2026, Volume 18, Issue 1
ISSN 1803-537X

bariéry, které odhalují rozdíl mezi čistě matematickou logikou a lidskou intuicí. Byly vymezeny tři roviny možných selhání:

A. Preference matematické abstrakce před heuristickým vylučováním

Modely AI mají tendenci převádět logické úlohy na soustavy rovnic či komplexní abstraktní struktury. Příkladem je úloha č. 13 (obrázek 1).

13. Petr do krajních políček napsal čísla 6 a 3. Do dalších pěti políček doplnil zbývající čísla od 1 do 7 tak, že součet čísel v každém kroužku je 10. Které číslo napsal Petr do políčka s otazníkem?



(A) 1 (B) 2 (C) 4 (D) 5 (E) 7

Obrázek 1 Zadání úlohy č. 13

Při řešení této úlohy se AI1 opakovaně a neúspěšně snažila o nalezení univerzálního matematického vzorce. Naproti tomu dětský přístup je přirozeně heuristický – pracuje s omezenou množinou přípustných prvků a využívá strategii postupného dosazování a vyřazování možností. Pro AI je často náročné opustit exaktní výpočetní cestu a přejít k intuitivnímu „pasování" prvků do vymezeného rámce.

B. Absence fyzikální intuice a hmatové zkušenosti

Úlohy založené na prostorové manipulaci odhalují absenci tzv. „embodied cognition" (vtěleného poznání). Příkladem je úloha č. 14 (obrázek 2).

14. Julie chce z dílků skládačky na obrázku sestavit housenku. Housenka má mít oba konce (s obličejem a banánem) a mezi nimi 1, 2 nebo 3 další dílky. Jaký největší počet různých housenek může Julie sestavit, aniž by dílky otáčela?

(A) 3 (B) 4 (C) 5 (D) 6 (E) 7

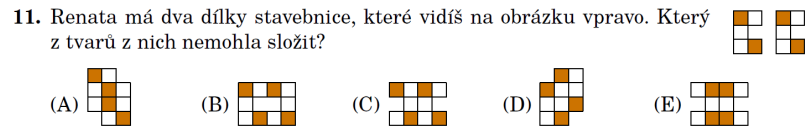
Obrázek 2 Zadání úlohy č. 14

Zatímco dítě disponuje zkušeností z fyzického světa – ví, že hmota je neprostupná a že tvary mají své mechanické limity – AI musí tyto koncepty

simulovat skrze logické výroky. Pro AI je obtížné nepravidelný tvar „dílku" jako fyzickou překážku, která blokuje pohyb; vnímá je pouze jako geometrická data, což vede k podcenění prostorových omezení, která jsou pro člověka samozřejmá

C. Kontextuální vizuální interpretace vs. digitální analýza

AI interpretuje grafické zadání jako soubor datových bodů, nikoliv jako scénu s určitou konvencí. Vizuální zkratky, které lidské oko dekóduje okamžitě (například to, že kružnice na kraji řady logicky zahrnuje méně políček než ty uprostřed – úloha č. 13, Obrázek 1), mohou být pro AI zdrojem nejednoznačnosti. Chybná interpretace grafického zadání úlohy byla indikovány např. i u úlohy č. 11 (Obrázek 3).



Obrázek 3 Zadání úlohy č. 11

AI nevnímala obrázek v zadání jako 2 samostatné dílky a úlohu jako mentálně manipulativní. Naopak, dle svého vyjádření, hledala opakovaně invarianty obrázku, případně primárně vnímala mřížovou strukturu bodů na obrázku.

Uvedené příklady ilustrují, že AI má tendenci preferovat formálně symetrické a matematicky elegantní modely, i v situacích, kdy zadání pracuje s asymetrií, implicitními pravidly nebo vizuálními zkratkami. Chyba zde nespočívá v samotném výpočtu, ale v porozumění kontextu a ve schopnosti opustit původně zvolený, avšak nevhodný interpretační rámec. Schopnost dětí provést „vizuální syntézu" – tedy pochopit implicitní pravidla z obrázku bez jejich explicitního vypsání v textu – zůstává pro umělou inteligenci často dosud výzvou.

5 Diskuse

Výsledky analýzy lze interpretovat v rámci konceptu vtěleného poznání (embodied cognition), podle něhož je lidské myšlení úzce propojeno s tělesnou a smyslovou zkušeností (Shapiro & Spaulding, 2025; Wilson, 2002). Žáci mladšího školního

1/2026, Volume 18, Issue 1
ISSN 1803-537X

věku při řešení geometrických a logicko-vizuálních úloh nevycházejí pouze z abstraktních matematických vztahů, ale zapojují zkušenosti získané fyzickou interakcí s okolním světem. Tyto zkušenosti jim umožňují rychle odhadnout, které varianty řešení jsou realistické a které lze okamžitě vyloučit. Jak upozorňují Lakoff a Núñez (2000), matematické myšlení je u člověka hluboce zakotveno v tělesné zkušenosti. Umělá inteligence naproti tomu pracuje s úlohami jako s formálně definovanými problémy, které je třeba převést do symbolického nebo algebraického jazyka, který však nemusí odpovídat skutečnému kontextu zadání. V důsledku toho může přehlížet význam vizuálních konvencí, implicitních omezení nebo „samozřejmých" pravidel, která nejsou v zadání explicitně formulována, ale jsou pro lidského řešitele intuitivně zřejmá.

Z didaktického hlediska se tak ukazuje, že generativní AI může být cenným nástrojem pro analýzu řešitelských strategií a chyb, nikoli však náhradou žákovského uvažování. Právě konfrontace „příliš analytického" přístupu AI s intuitivními a heuristickými strategiemi dětí může představovat významný zdroj pro rozvoj metakognice a kritického myšlení ve výuce matematiky na 1. stupni základní školy.

Předložená studie má několik omezení. Výzkum byl realizován na omezeném souboru úloh jedné soutěže a výsledky proto nelze bez dalšího zobecňovat na všechny typy matematických úloh. Současně byly analyzovány pouze tři konkrétní systémy generativní AI, jejichž vlastnosti se mohou v čase měnit v důsledku aktualizací modelů. Dalším omezením je použití zero-shot promptingu, který sice odpovídá běžnému uživatelskému využití, avšak nemusí odrážet maximální potenciál jednotlivých systémů. Budoucí výzkum by měl zahrnout širší spektrum úloh, více AI modelů a také srovnání s řešitelskými strategiemi skutečných žáků.

6 Závěr

Předložená studie se zaměřila na analýzu chyb v řešeních matematických úloh generovaných vybranými systémy umělé inteligence, konkrétně v úlohách kategorie Klokánek soutěže Matematický klokan. Výsledky ukázaly, že jednotlivé AI systémy se v úspěšnosti řešení úloh statisticky významně lišily. Zároveň se potvrdilo, že chyby se výrazně soustředily do úloh s geometrickým a logickým kontextem, které vyžadují práci s vizuálním uspořádáním, implicitními pravidly zadání a prostorovou představivostí. Analýza vybraných úloh dále ukázala, že chybné řešení často nevzniká v důsledku izolované numerické chyby, ale jako

důsledek přetrvávání nevhodně zvolené interpretační strategie, která může přetrvat i při opakované interakci s AI.

Acknowledgements

Příspěvek vznikl v rámci realizace projektu „Inovativní přístup k přípravě budoucích učitelů matematiky v éře umělé inteligence“ č. projektu: IGA_PdF_2025_034, realizovaného na Katedře matematiky Pedagogické fakulty Univerzity Palackého v Olomouci.

7 Literatura

Bártek, K., Bártková, E., Mrkvan, K., & Nocar, D. (2025). Možnosti a limity LLM v geometrické přípravě budoucích učitelů 1. stupně základních škol. *Elementary Mathematics Education Journal, 7*(2). 89–103. https://emejournal.upol.cz/Issues/Vol7No2/Vol7No2_Bartek-et-al.pdf

Boye, J., & Moell, B. (2025). Large language models and mathematical reasoning failures. arXiv. https://arxiv.org/abs/2502.11574

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., & Steinhardt, J. (2021). Measuring Mathematical Problem Solving With the MATH Dataset. arXiv. https://arxiv.org/abs/2103.03874

Kambhampati, S. (2024). Can large language models reason and plan? *Annals of the New York Academy of Sciences, 1534*(1), 15–18. https://doi.org/10.1111/nyas.15125

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books. https://pages.ucsd.edu/~rnunez/COGS252_Readings/Preface_Intro.PDF

Shapiro, L., & Spaulding, S. (2025) *Embodied Cognition*. The Stanford Encyclopedia of Philosophy (Summer 2025 Edition). https://plato.stanford.edu/archives/sum2025/entries/embodied-cognition/

Schoenfeld, A. H. (2016). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics (Reprint). *Journal of Education*, *196*(2), 1–38.

Strohmaier, A. R., Van Dooren, W., Seßler, K., Greer, B., & Verschaffel, L. (2025). Large language models don't make sense of word problems: A scoping review from a mathematics education perspective. arXiv. https://arxiv.org/abs/2506.24006

UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Xu, W., Wang, J., Wang, W., Chen, Z., Zhou, W., Yang, A., Lu, L., Li, H., Wang, X., Zhu, X., Wang, W., Dai, J., & Zhu, J. (2025). VisuLogic: A Benchmark for Evaluating Visual Reasoning in Multi-modal Large Language Models. arXiv. https://arxiv.org/abs/2504.15279

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). *Chain-of-thought prompting elicits reasoning in large models.* Advances in Neural Information Processing Systems, 35, 24824–24837.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636. https://doi.org/10.3758/BF03196322