

Dynamický Turingův test: Adaptace rozhodovacích strategií při detekci impresionistických maleb generovaných umělou inteligencí

Nikita Kostka, Ondřej Vadinský^[0000–0002–0910–3140]

Department of Information and Knowledge Engineering, Prague University of Economics and Business
Náměstí Winstona Churchilla 4, 130 67 Praha 3, Czech Republic
Email: nikita.kostka@gmail.com

Abstrakt

Príspevek navrhuje dynamickou variantu grafického Turingova Testu, která umožňuje sledovat adaptaci respondentů mezi třemi koly na základě zpětné vazby. V úvodním experimentálním šetření byla sledována schopnost vysokoškolských studentů ($n = 9$) rozlišovat mezi impresionistickými malbami a obrazy generovanými modelem Midjourney (3×5 dvojic). Výsledky ukazují statisticky slabě signifikantní nárůst úspěšnosti mezi prvním a třetím kolem (z 57,7 % na 84,4 %). Pro zhodnocení psychometrických kvalit testu byla využita analýza dle principů teorie odpovědi na položku. Kvalitativní analýza odhalila tři dominantní strategie: posuzování vizuální složitosti (úspěšnost 84,5 %), řízení se uměleckým principem (úspěšnost 84,6 %) a zaměření na detaily (úspěšnost 51,6 %).

1 Úvod

Generativní umělá inteligence dnes dokáže stále věrohodněji napodobovat lidskou tvorbu, a proto je důležité zkoumat, nakolik ji lidé ještě zvládají rozpoznat. Problematikou rozlišení mezi člověkem a strojem se zabýval již britský matematik Alan Turing (1950), který navrhl experiment umožňující tuto skutečnost testovat. Tento experiment, později nazvaný Turingův test, však vznikl v době, kdy byl vývoj umělé inteligence teprve v počátcích. V kontextu dnešních technologií, zejména multimodálních modelů viz např. (Ramesh a spol., 2022; Schellaert a spol., 2023), není již původní čistě textová podoba testu postačující. Již Harnad (1989) a po něm Schweizer (2012) navrhovali rozšíření testu, aby zahrnul plné spektrum (lidského, resp. druhově specifického) inteligentního chování, byť i jím stále šlo hlavně o lepší uchopení porozumění jazyku. Pro tento příspěvek je pak významnou inspirací především převedení Turingova testu do vizuální domény, které navrhl Michael McGuigan (2006). Turingův test i jeho dosavadní modifikace však uvažují statické měření úspěšnosti a opomíjejí proces učení, který je také podstatnou částí (nejen) lidské kognice. Předložený příspěvek proto rozšiřuje Turingův test o dynamický princip, a přenáší jej z původní textové domény do domény obrazové.

Dále příspěvek prezentuje výsledky úvodního experimentálního ověření navrženého testu. Toto ověření sleduje dva hlavní cíle: validaci samotného metodického návrhu a následný sběr i vyhodnocení kvalitativních dat od respondentů. Právě analýza kvalitativních výsledků umožňuje identifikovat a komparovat rozhodovací strategie účastníků v jednotlivých fázích testu. Tento přístup otevírá prostor pro detailní zkoumání procesu adaptace respondentů a sledování dynamických změn v jejich úsudku v průběhu celého experimentu.

Příspěvek nejprve v sekci 2 vysvětlí teoretická východiska. Proces návrhu dynamického Turingova testu pro rozpoznávání generovaných impresionistických maleb představí sekce 3. Provedený demonstrační experiment popíše sekce 4. Jeho výsledky ukáže sekce 5. Diskuzi těchto výsledků se bude věnovat sekce 6. Příspěvek bude zakončen v sekci 7.

2 Teoretická východiska návrhu

Sekce shrnuje hlavní oblasti, na jejichž základě byl rozšířený Turingův test navržen. Technologický základ pro generování věrohodných obrazových stimulů tvoří modely generativní umělé inteligence založené na principu difúze, konkrétně model Midjourney. Teoretický základ představuje samotný Turingův test a jeho limity, přičemž východiskem úprav pro sledování kognitivní adaptace jsou principy dynamického testování.

Difúzní modely stručně představí sekce 2.1. Turingovu testu a jeho limitům se bude věnovat sekce 2.2. Principy dynamických testů vysvětlí sekce 2.3.

2.1 Difúzní modely

Difúzní modely představují v současnosti jednu z nejprogresivnějších architektur v oblasti generativní umělé inteligence zaměřené na tvorbu obrazových dat. Jejich princip vychází z pravděpodobnostního modelování a teorie stochastických diferenciálních rovnic. Model se v rámci trénování učí dvěma základními procesům: dopředné difúzi, která postupně transformuje reálná data na šum, a následnému reverznímu procesu, při kterém dochází k rekonstrukci původních dat z náhodného šumu (Park a spol., 2023).

Proces postupné degradace obrazu (zašumování) lze v diskretním čase $t \in \{1, \dots, T\}$ popsat rovnicí:

$$x_t = \sqrt{1 - \beta_t} x_{t-1} + \sqrt{\beta_t} \epsilon_t \quad (1)$$

kde ϵ_t představuje nezávislé standardní Gaussovo rozdělení a skalární parametr $\beta_t \in (0, 1)$ určuje množství šumu přidaného v každém kroku. Pro další modelování se zavádí proměnná $\alpha_t = 1 - \beta_t$, vyjadřující míru zachované informace. Kumulativní vliv procesu od počátečního stavu x_0 do času t je pak definován jako:

$$\bar{\alpha}_t = \prod_{i=1}^t \alpha_i \quad (2)$$

Cílem trénovaného modelu je predikovat přidaný šum a umožnit tak inverzní transformaci. Odšumovací krok, který směřuje od čistého šumu x_T zpět k realistickému výstupu x_0 , je definován vztahem:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_q(t) z \quad (3)$$

kde $\epsilon_\theta(x_t, t)$ je neuronová síť odhadující šum v daném kroku a $\sigma_t z$ představuje stochastickou složku zajišťující rozmanitost generovaných výstupů. Pro detailnější rozbor difúzních modelů viz Higham a spol. (2023), či Ho a spol. (2020).

2.2 Turingův test

Turingův test, který je hlavní inspirací a zdrojem pro tento příspěvek, navrhl v roce 1950 britský matematik a logik Alan M. Turing ve svém článku *Computing machinery and intelligence* publikovaném v časopise *Mind*. Při experimentu jsou myšleni tři účastníci: A , B , C a hrají tzv. *imitační hru*. Jeden ze dvojice A , B je člověk, zatímco druhý je stroj. Oba komunikují s vyšetřovatelem C výhradně prostřednictvím textových zpráv (aby nebyl rozpoznatelný jejich vzhled ani chování) a snaží se ho přesvědčit, že právě oni jsou člověkem. Úkolem vyšetřovatele je na základě konverzace určit, kdo z účastníků je člověk a kdo stroj. To má vyšetřovatel určit tak, že se účastníkům A a B zadává dotazy a hodnotí jejich odpovědi dle toho, zda by takto dokázal odpovědět člověk (Turing, 1950).

Hlavní limit takto navrženého testu spočívá v jeho behaviorálním redukcionismu, kdy je inteligence posuzována výhradně na základě vnějších projevů v izolovaném prostředí, bez ohledu na vnitřní mentální procesy. Jak ilustruje Searlův experiment *Čínský pokoj*, úspěch v testu může reflektovat pouze syntaktickou dovednost manipulace se symboly, nikoliv skutečné sémantické porozumění nebo vědomí (Searle, 1980). Pro hlubší diskusi filosofických aspektů viz např. Tvrdý (2014). Dalším kritickým faktorem je subjektivita lidského pozorovatele a binární povaha výsledku, která neumožňuje hlubší analýzu kognitivních strategií,

jimiž systém k odpovědím dospěl (French, 2000). Pro diskuzi limitů Turingova testu jakožto testu (strojové) inteligence viz např. Hernández-Orallo (2000).

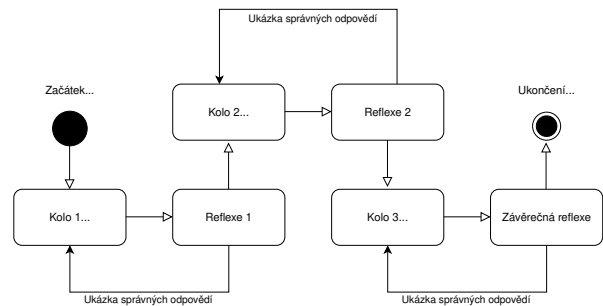
2.3 Dynamické testování

Dynamické testování představuje diagnostický přístup zaměřený na proces učení a kognitivní přizpůsobivost respondenta v čase. Na rozdíl od tradičních statických metod, které měří aktuální úroveň schopností jako neměnný stav, dynamické pojetí vychází z předpokladu, že se lidský výkon vyvíjí v interakci s prostředím a skrze poskytovanou intervenci. Klíčovým prvkem je interaktivní charakter experimentu: respondent dostává cílené instrukce, nápovědy či zpětnou vazbu, což mu umožňuje aktivně zdokonalovat své strategie a nově nabyté poznatky okamžitě aplikovat v opakovaných fázích testu (Grigorenko, 2009; Lidz, 2017). Pro širší představení dynamického testování viz Sternberg (2002).

3 Návrh dynamického Turingova testu

Tato sekce přechází od teoretických východisek k praktickému návrhu dynamického Turingova testu v doméně vizuálního umění. Hlavním cílem je vytvoření experimentálního rámce pro identifikaci a analýzu procesu kognitivní adaptace respondentů při rozpoznávání generovaných impresionistických maleb. Text popisuje celkovou strukturu testu a vysvětluje proces konstrukce jednotlivých testových položek.

Návrh předloženého dynamického Turingova testu představí sekce 3.1. Konstrukci položek se bude věnovat sekce 3.2.



Obr. 1: Diagram průběhu testu

3.1 Struktura testu

Navržená struktura přímo reflektuje principy dynamického testování popsané v sekci 2.3. Test je koncipován jako cyklický proces umožňující kognitivní adaptaci respondentů skrze průběžnou zpětnou vazbu. Celý experiment sestává ze tří kol, z nichž každé obsahuje pět položek. Jednu položku tvoří dvojice reálné impre-

sionistické malby a obrazu generovaného AI (pro bližší představení konstrukce položek viz sekci 3.2). Respondent je vyzván, aby v každé dvojici identifikoval generovaný obraz, popsal důvody své volby a zároveň uvedl hodnotu jistoty od 1 do 5 (kde 1 je nejnižší jistota a 5 nejvyšší). Zároveň je respondentovi měřen čas výběru. Všechny tyto hodnoty jsou následně zaznamenávány. Po dokončení každé pěti položek následuje fáze reflexe, v níž jsou respondentovi odhaleny správné odpovědi, čímž získává prostor pro korelaci svých poznatků a jejich aplikaci v následujících blocích.

3.2 Konstrukce položek testu

Metodika konstrukce testových položek spočívala v systematickém párování originálních děl s jejich generovanými ekvivalenty. Z digitálních archivů WikiArt¹ a Wikimedia Commons² bylo náhodně vybráno 15 reprezentativních impresionistických děl, každé od odlišného autora.

K tvorbě obrazů byl využit nástroj Midjourney³. Proces generování využíval kombinaci obrazové reference a textového popisu. Původní dílo bylo nahráno jako stylová reference, k níž byl připojen anglický prompt uvozený termínem *impressionism* a následovaný stručným popisem hlavního motivu malby.



Obr. 2: Ukázková položka testu: A1 – reálné dílo, A2 – generovaný obraz

Na Obr. 2 je zachycena ukázková dvojice: reálný obraz *William Merritt Chase – „Shinnecock Hills“* (A1) a jeho generovaný protějšek (A2) vytvořený na základě promptu *impressionism, road in fields of dry grass*.

4 Demonstrační experiment

Výše představený návrh testu byl následně podroben úvodnímu experimentálnímu ověření. Sběr dat probíhal v říjnu 2025 a testování se zúčastnilo 9 respondentů.

¹Archiv je přístupný na <https://www.wikiart.org>.

²Archiv je přístupný na <https://commons.wikimedia.org>.

³Nástroj je přístupný skrze rozhraní na <https://www.midjourney.com>.

Respondenty blíže představí sekce 4.1. Sekce 4.2 upřesní nástroje, jež byly při testování použité.

4.1 Respondenti

Respondenty tvořilo devět studentů z osmi českých vysokých škol v různých úrovních studia a studujících různé obory. Byli zastoupeni respondenti s odlišnými technickými znalostmi a dovednostmi, s různými zkušenostmi a postoji k využívání generativní umělé inteligence a také s rozdílnými vztahy k výtvarnému umění.

Aby byly podmínky testu metodické a srovnatelné, dodržoval výzkum doporučení Hendl (2005); Urbánek (2011) týkající se etických, administrativních a organizačních zásad. Mezi tyto principy patřilo zajištění standardizovaných podmínek testování, informovaný souhlas respondentů, srozumitelnost zadání, neutrální a objektivní pokyny a respekt k individualitě respondentů.

4.2 Použité nástroje při testování

Sběr dat probíhal formou audiozáznamu, kvantitativní tabulky a doplňujících textových poznámek administrátora. K realizaci byla využita dvě mobilní zařízení: notebook administrátora pro průběžnou evidenci údajů a měření času a tablet pro respondenta, který sloužil k interaktivní prezentaci obrazových stimulů a pořízení zvukové nahrávky. Jednotlivé testové položky byly respondentům prezentovány ve formě obrázků s možností přiblížení pro lepší analýzu detailů.

5 Výsledky

Sekce představí výsledky získané v rámci demonstračního experimentu. Kvantitativní výsledky byly analyzovány za pomoci statistických metod, kvalitativní výsledky za pomoci kódování jednotlivých odpovědí v prostředí programu QDA Miner.

Kvantitativní výsledky představí sekce 5.1. Kvantitativní výsledky představí sekce 5.2. Bližší pohled na vývoj a použití strategií mezi koly ukáže sekce 5.3. Na závěr sekce 5.4 představí provedenou analýzu teorie odpovědi na položku.

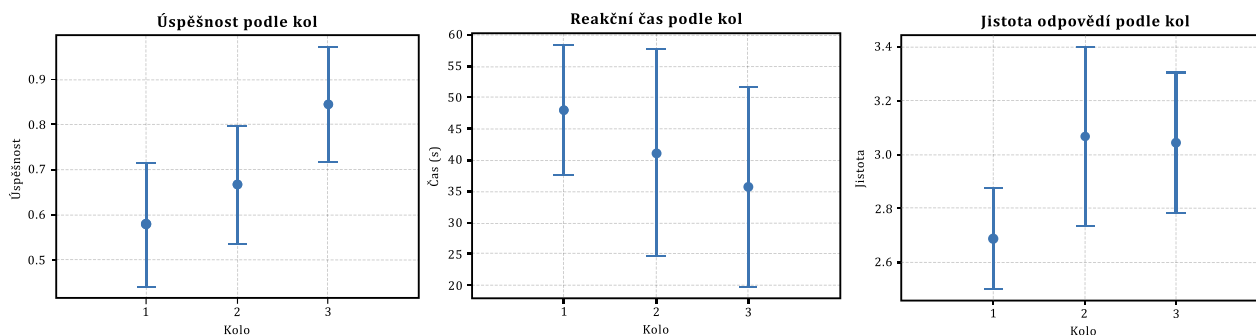
5.1 Kvantitativní výsledky

V rámci kvantitativních dat bylo nasbíráno 405 údajů ve třech sledovaných proměnných:

Správnost: booleovská hodnota, zda respondent vybral správně generovaný obrázek;

Čas: délka výběru v sekundách;

Jistota: subjektivní míra jistoty na stupnici od jedné do pěti (1 = nejmenší jistota výběru; 5 = největší).



Obr. 3: Průměrné hodnoty úspěšnosti, reakčního času a subjektivní jistoty odpovědí v jednotlivých kolech testu včetně 95% intervalů spolehlivosti.

Proměnná	Kolo 1	Kolo 2	Kolo 3
Průměrná správnost (%)	57,70	66,66	84,44
Průměrná jistota	2,68	3,06	3,04
Průměrný čas (s)	47,95	41,17	35,68

Tab. 1: Průměrné hodnoty proměnných v jednotlivých kolech.

Jak lze pozorovat ze zprůměrovaných dat v Tab. 1, průměrná správnost výběru se mezi koly postupně zvyšovala, zatímco průměrný čas výběru se snižoval. Je zde tedy přítomná jistá tendence ke zlepšování, ovšem tu se podařilo statisticky prokázat jen částečně. Jediné signifikantní zlepšení bylo v úspěšnosti mezi prvním a třetím kolem, kde dosáhl párový t -test výsledku $p = 0,0222$. Průměrná jistota mezi prvním a druhým kolem stoupla a poté se spíše ustálila.

Pro lepší představu jsou výsledky vyznačeny v grafech s 95% intervalem spolehlivosti na Obr. 3. Velký rozsah intervalů spolehlivosti je dán nízkým počtem subjektů ($n = 9$) a v některých případech zřejmě i vyšší variabilitou v jejich odpovědích. Pozorovatelné překryvy intervalů spolehlivosti naznačují, že ve většině případů nejsou průměrné hodnoty sledovaných proměnných signifikantně odlišné, což je v souladu s výsledky provedeného párového t -testem.

5.2 Kvalitativní výsledky

Jak již bylo zmíněno v úvodu této sekce, kvalitativní výsledky byly analyzovány za pomoci metody kódování. Při ní se jednotlivé odpovědi respondentů s podobným smyslem označí společným kódem, pod kterým jsou seskupeny. Výsledné rozdělení ukazuje Tab. 2.

Již na první pohled je zřejmá nejčastěji používaná strategie, kterou je pozorování nepřírozené složitosti generovaných obrazů. Tato strategie také patří k jedné z neúspěšnějších. Dalšími dvěma nejčastěji použitými strategiemi je pozorování detailů a analýza obrazů na základě uměleckého principu.

Kód	Počet výskytů	Úspěšné určení	Úspěšnost (%)
Plátno	5	3	84,5
Detaily	31	16	51,6
Složitost	58	49	84,5
Odras vody	11	6	54,5
Pocitově	21	11	52,4
Umělecký princip	26	22	84,6
Zkušenosti s AI	6	4	66,7
Bez uvedení důvodu	17	12	70,6

Tab. 2: Kvalitativní výsledky – kódy strategií.

5.3 Vývoj strategií mezi koly

Mnohem zajímavější je ovšem pohled na celkový vývoj použití jednotlivých strategií a jejich aplikaci následující reflexi předešlého kola. Tam je nejvíce patrné, jak v závislosti na zpětné vazbě dokázali respondenti upravovat použití jednotlivých strategií. Níže je rozebrán vývoj tří nejčastějších strategií představených v minulé sekci.

Nejlepší ukázkou správné adaptace je vývoj strategie pozorování umělé složitosti generovaných maleb. Ta byla v prvním kole použita ve 12 případech, ovšem pouze u 6 z nich vedla k úspěšnému rozlišení. Dosáhla tedy úspěšnosti 50 %, což je stejné jako náhodné hádání. Poté proběhla první reflexe a respondenti měli možnost zhodnotit své dosavadní výsledky. Ve druhém kole byla tato strategie využita již v 17 případech, ze kterých v 16 případech pomohla ke korektnímu určení a dosáhla tedy úspěšnosti 94 %. Je tedy vidět relativně slabý nárůst v množství výskytů, ale znatelný nárůst ve správnosti určení. Ve třetím kole byla poté tato strategie využita ve 29 případech, z čehož ve 27 vedla ke správnému určení a dosáhla tak úspěšnosti 93 %. Lze tak pozorovat, že zatímco úspěšnost zůstala stále extrémně vysoká, značně se také zvýšilo nasazení této strategie. Vývoj využití této strategie by tak mohl naznačovat, že ji účastníci po první reflexi začali vnímat jako účinnější a začali na ni více spoléhat. Značné zlepšení mezi prvním a druhým kolem také může značit, že se respondenti naučili používat ji efektivněji v určitých případech.

Opačný případ lze pozorovat u strategie pozorování detailů. Tato strategie byla v prvním kole na

sazena v 17 případech a k úspěšnému odhalení generované malby vedla v 11 z nich. Dosáhla tedy 64% úspěšnosti. Ve druhém kole byla využita v 11 případech, ovšem pouze ve 4 z nich úspěšně a dosáhla tak úspěšnosti pouze 36 %. Lze tedy pozorovat, že se samotné nasazování strategie snížilo, ale ještě více se snížila její úspěšnost. Ve třetím kole byla využita již pouze ve třech případech a pouze v jednom z nich vedla ke správnému určení. Tento vývoj naznačuje, že v průběhu testu začali respondenti upřednostňovat jiné a účinnější strategie a od této pomalu upouštěli.

Zajímavé také je, že v prvním kole se jednalo zcela o nejpoužívanější strategii, což poukazuje na obecný předsudek respondentů vůči umělé inteligenci, že chyby v generování jsou nejlépe patrné v chybách detailů. Nízká úspěšnost strategie ovšem může znamenat, že generativní modely již neprodukují zjevné chyby v detailech, které by bylo možné snadno odhalit, či že se nejedná o vhodný znak generování specificky u impresionistických obrazů, kde umělci často nevěnovali detailům přílišnou pozornost a snažili se hlavně o působení obrazu jako celku.

Třetí nejčastější strategie, tedy řízení se uměleckým principem má vývoj relativně stálý. V prvním kole byla tato strategie použita v 8 případech, z čehož v 6 efektivně. Měla tedy 75% úspěšnost. Ve druhém kole byla využita opět v 8 případech, ale v 7 vedla ke správnému určení. Ve třetím kole byla využita v 10 případech a v 9 z nich efektivně. Lze tedy pozorovat relativně stálý počet aplikace strategie a také stálou, relativně vysokou úspěšnost. Tento výsledek by mohl značit, že se jedná o spolehlivou strategii nejlépe aplikovatelnou u určitých položek, kde jsou patrné stylové prvky.

Tato analýza jednotlivých strategií ukazuje, že zaměření na změny mezi koly testu přináší nový a hlubší pohled do adaptací jednotlivých respondentů a mělo by smysl je sledovat více. Ovšem je důležité upozornit, že v rámci testování nebyl pro respondenty limitován počet strategií při výběru jedné položky. Tedy u jednoho rozhodnutí mohlo být použito více strategií zároveň.

5.4 Teorie odpovědi na položku

Pro zhodnocení položek testu byla využita analýza za pomoci teorie odpovědi na položku, jak popisují Allen a Yen (2001). Ta odhalila značné limity v některých položkách i struktuře testu. Výsledky ukazují, že test byl pro respondenty spíše jednodušší (průměrná p-hodnota 0,69) a celková průměrná diskriminační korelace (0,092) signalizuje nižší strukturní stabilitu testu. Ovšem interpretace obtížnosti je částečně ovlivněna fixním řazením položek. Vysoká úspěšnost v pozdějších fázích může být důsledkem procesu učení, což je v souladu s dynamickou povahou testu.

Podářilo se však určit položky jak se silnou, tak i se slabou výpovědní hodnotou. U některých položek

byla korelace záporná, což značí nekonzistenci s celkovým výkonem. Slabší respondenti na ně tedy odpovídali lépe než ti silnější.

Analýza IRT tak poskytuje podklad pro budoucí optimalizaci výběru a pořadí stimulů.

6 Diskuse výsledků

Tato sekce interpretuje získaná data, hodnotí metodické limity testu a navrhuje směry pro navazující výzkum. Výsledky jsou nejdříve zasazeny do širšího kontextu v sekci 6.1. Kritická reflexe limitů následuje v sekci 6.2. Možnosti dalšího rozvoje rozebírá sekce 6.3.

6.1 Interpretace hlavních zjištění

Z kvantitativních výsledků se statisticky podložené zlepšení projevilo pouze u úspěšnosti v porovnání prvního a třetího kola. U zbývajících hodnot, jako je průměrný čas výběru a subjektivní jistota, nelze kvůli nízkému počtu participantů činit definitivní závěry. Přesto lze mezi jednotlivými koly pozorovat tendenci ke snižování časové latence a postupné stabilizaci jistoty respondentů. Kvalitativní analýza naopak jasně ukázala, že se v průběhu testu stala dominantní strategie zaměřená na umělou složitost obrazu, přičemž respondenti v její prospěch opouštěli méně účinné postupy. Spojením těchto zjištění lze vyvodit, že respondenti v průběhu testu aktivně vyhodnocují své strategie, což jim pomáhá k úspěšnějšímu rozlišování mezi reálnou a generovanou tvorbou. Návrh testu tak i přes své limity nabízí funkční metodický rámec pro sledování kognitivní adaptace.

6.2 Reflexe a limity

Navržený test má v nynější podobě řadu limitů. Nejzákladnějším faktorem je malý počet respondentů ($n = 9$), což je způsobeno časovou náročností hloubkových rozhovorů. Při větším vzorku by bylo možné dosáhnout robustnějších výsledků vhodnějších pro statistické ověření. Další limity vyplývají z analýzy IRT, která odhalila nevyrovnanou obtížnost a diskriminační schopnost některých testových dvojic, což snižuje ucelenost testu.

Zásadním omezením je také výhradní zaměření na impresionistické malby. Toto rozhodnutí sice zjednodušilo analýzu, ale limituje generalizaci výsledků na jiné umělecké styly či doménu fotografie. Identifikovaná strategie pozorování složitosti může být specifická právě pro impresionismus, kde modely generativní AI zatím nedokáží plně napodobit charakteristický rukopis. V neposlední řadě představuje limit i neustálý technický vývoj AI. Je pravděpodobné, že s nástupem dokonalejších modelů budou současné identifikační strategie vyžadovat revizi.

6.3 Možnosti dalšího výzkumu

Budoucí výzkum by se měl zaměřit na eliminaci zmíněných limitů. Přírozenou cestou je aplikace testu na další umělecké směry, což umožní analyzovat proměnlivost strategií v závislosti na vizuálním charakteru děl. Zajímavý vhled by mohlo přinést také zkoumání vlivu informovanosti respondenta, tedy zda a jak se změní úspěšnost výběru, pokud je účastníkům předem vysvětlen princip fungování generativních modelů.

Zásadní prostor pro rozvoj představuje hlubší analýza individuálních strategií. Dosavadní data naznačují, že detailní sledování toho, jak se konkrétní strategie u jednotlivých respondentů v čase utvrzují či mění, nabízí nejvyšší výpovědní hodnotu o procesu lidského učení. V metodologické rovině se dále nabízí randomizace pořadí položek pro eliminaci únavového efektu nebo vytvoření různých položek, které budou nejprve samostatně zhodnoceny pro co nejvyšší rozlišovací schopnost.

V neposlední řadě by bylo přínosné srovnání výkonu laiků s experty z oblasti dějin umění a vývojářů AI, kteří by mohli k rozlišování přistupovat na základě odlišných zdůvodnění výběru.

7 Závěr

Tento příspěvek představil modifikaci Turingova testu převedením do obrazové domény a jeho rozšířením o dynamický princip. Tato úprava umožnila sledovat změny v rozhodovacím procesu respondentů a analyzovat, jakým způsobem přizpůsobují své strategie na základě získané zpětné vazby.

Kvantitativní data potvrdila statisticky významné zlepšení úspěšnosti mezi počátkem a koncem testu. Kvalitativní analýza identifikovala klíčové strategie, z nichž neefektivnější se ukázalo posuzování kompoziční složitosti.

I přes popsané limity v podobě malého výzkumného souboru a úzkého stylového zaměření práce potvrzuje, že dynamický přístup k testování poskytuje hlubší vhled do interakce člověka s obsahem generovaným umělou inteligencí. Předložený koncept testu může sloužit jako základ pro další zkoumání hranic mezi lidskou a strojovou tvorbou v neustále se vyvíjejícím digitálním prostředí.

Reference

Allen, M. J. a Yen, W. M. (2001). *Introduction to Measurement Theory*. Waveland Press.

French, R. M. (2000). The Turing Test: the first 50 years. *Trends in Cognitive Sciences*, 4(3):115–122.

Grigorenko, E. L. (2009). Dynamic Assessment and Response to Intervention. *Journal of Learning Disabilities*, 42(2):111–132.

Harnad, S. (1989). Minds, machines and Searle. *Journal of Experimental & Theoretical Artificial Intelligence*, 1(1):5–25.

Hendl, J. (2005). *Kvalitativní výzkum: základní metody a aplikace*. Portál.

Hernández-Orallo, J. (2000). Beyond the Turing test. *Journal of Logic, Language and Information*, 9(4):447–466.

Higham, C. F., Higham, D. J. a Grindrod, P. (2023). Diffusion Models for Generative Artificial Intelligence: An Introduction for Applied Mathematicians.

Ho, J., Jain, A. a Abbeel, P. (2020). Denoising Diffusion Probabilistic Models.

Lidz, C. (2017). Dynamic Assessment: Thinking Assessment for the Thinking Teacher I Carol S. Lidz. *The Thinking Teacher*, Volume 1.

McGuigan, M. (2006). Graphics Turing Test.

Park, Y.-H., Kwon, M., Choi, J., Jo, J. a Uh, Y. (2023). Understanding the Latent Space of Diffusion Models through the Lens of Riemannian Geometry.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. a Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents.

Schellaert, W., Martínez-Plumed, F., Vold, K., Burden, J., Carases, P. A. M., Loe, B. S., Reichart, R., Héigeartaigh, S. Ó., Korhonen, A. a Hernández-Orallo, J. (2023). Your prompt is my command: On assessing the human-centred generality of multimodal models. *Journal of Artificial Intelligence Research*, 2023(77):377–394.

Schweizer, P. (2012). Externalist foundations of a truly total Turing test, the. *Minds and Machines*, 22(3):191–212.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, (3):417–457.

Sternberg, R. J. (2002). *Dynamic testing: the nature and measurement of learning potential*. Cambridge, UK ; New York : Cambridge University Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.

Tvrđý, F. (2014). *Turingův Test: Filozofické Aspekty Umělé Inteligence*. Togga, Praha, 1. vyd.

Urbánek, T. (2011). *Psychometrika: měření v psychologii*. Portál.