

Multimodální AI: rozpoznávání změn ve scéně

Vytvoříte vlastní obrazový test a na stejných fotografiích porovnáte pět různých modelů. Důležité je poctivě měření a vysvětlení výsledků, ne vítězství konkrétní AI.

Práce jednotlivce. Měření: 1. 10.–15. 11. 2026.

Odevzdání: 15. 11. 2026. Odkaz na odevzdávací zveřejní vyučující.

12+

vlastních scén
na studenta

48+

fotografií
4 na scénu

5

dvojic
na scénu

5+

různých modelů /
checkpointů

300+

odpovědí
v hlavní části*

Jedna sada = jedna scéna ve čtyřech stavech. Nejméně dvě jednoznačně měnící se vlastnosti a jedna stabilní vlastnost. Snímky nemusí tvořit příběh; každý test porovnává samostatnou dvojici.

Inspirace: čtyři stavy jedné scény



1 / Výchozí stav. Flakon nápisem k nám, kartáč štětinami dolů.



2 / Jiné natočení. Flakon bokem, kartáč a krabička pootočené.



3 / Více změn. Taška není vidět, kartáč a hrnek mají jiný stav.



4 / Převrácení a otočení. Flakon hladkou stranou, kartáč štětinami nahoru.

Učitelská ukázka parapetu, nikoli vaše odevzdávaná data. Snímky jsou celé, bez ořezu. Pro vlastní scény držte fotoaparát i záběr na stejném místě.

* Minimum plánujte jako $12 \times 5 \times 5 = 300$ dokončených odpovědí. Doložené technické výpadky neposuzujte jako chyby AI; odchylku od dokončeného minima musí schválit vyučující (str. 4). Retry není nová odpověď do minima.

Co připravit před měřením

1 / Vlastní scény

Nejméně 12 různých scén. V každé čtyři fotografie 01–04. Jasně změny polohy, orientace, přítomnosti nebo stavu; kreativita nesmí znamenat nečitelnost.

2 / Jeden společný klíč

Do ground_truth.csv zapište všechny ověřitelné změny a stálost ve čtyřech sloupcích. Tabulka je současně popisem; dalších 48 slohových popisů nepíšete.

3 / Dva spolužáci

Každou scénu posoudí dva různí lidé mimo autora: nejprve snímky bez klíče, potom klíč. Zapište ID, čas a schválení do validation.csv. Nejasnosti vyřešte před AI.

4 / Zámek a plán

Vyplňte modely a project.json, spusťte kontrolu, vytvořte plan.csv a freeze.json. Po prvním pokusu už neměňte klíč podle odpovědí modelů.

Jak vybrat pět skutečně různých modelů

Model musí přijmout **dva obrázky a text v jednom požadavku**. Lze použít web, API i lokální model a kombinovat poskytovatele. Placená služba není podmínkou. Pro hlavní část použijte výchozí nastavení; nutné odchylky zapište.

Počítá se jako další model	Nepočítá se jako další model
Jiný základní model / checkpoint s odlišnými vahami. Dva různé modely jedné rodiny jsou přípustné.	Stejně váhy s jiným reasoning režimem, teplotou, kvantizací nebo jiným poskytovatelem.
Např. Qwen3.8 Flash Next a Qwen3.8 27B.	Stejný model v režimu „default“ a „thinking“.

Do checkpoint_identity zapište společnou identitu základních vah, nikoli vlastní označení režimu. Skutečnou odlišnost ověřuje člověk; validátor pouze odhalí shodné zapsané identity.

Co zaznamenat v models.csv

- Poskytovatele, službu, zobrazované jméno, přesné zveřejněné model ID, rodinu a identitu checkpointu.
 - Rozhraní, režim a nastavení. Pro lokální běh také kvantizaci, software a hardware.
 - Skutečné datum prvního a posledního pokusu. Neveřejnou verzi nevymýšlejte: prázdné exact_model_id vysvětlíte v notes.
- Stejný vstup, ne domněnka o stejném backendu.** Všem modelům dejte stejné soubory, pořadí a P1. Nastavení jednoho modelu po celou hlavní část neměňte. Rozdíly rozhraní či neveřejné výchozí hodnoty přiznejte jako omezení.

Fotografie musí být vlastní, bez AI úprav, osobních dokumentů, hesel a identifikovatelných lidí bez oprávnění. Každý změnový pár 1-2, 1-3, 1-4 a 4-1 musí mít alespoň jednu změnu v klíči. Stejný stav zapisujte vždy stejnými slovy, aby šlo tabulky automaticky porovnat.

Modely a přístup k e-INFRA

Doporučená pětice pro tento projekt

Model	Konkrétní API označení
Gemma 4	gemma4
Kimi K3	kimi-k3
Mistral Medium 3.5	mistral-medium-3.5
Qwen3.8 Flash Next	qwen3.8-flash-next
Qwen3.8 27B	qwen3.8-27b

Dokumentace CERIT-SC uvádí tuto pětici jako multimodální. Ověřeno 28. 9. 2026; nejde o garanci dostupnosti během semestru. Jiné vhodné modely jsou povoleny. [Zdroj: Chat AI — modely a přístup.](#)

1 / Přístup a přihlášení

Připravte si aktivní účet MetaCentra; pokud jej nemáte, vyřídte registraci školní identitou s předstihem. Po aktivaci otevřete [chat.ai.e-infra.cz](#) a přihlaste se přes e-INFRA CZ.

[Registrace MetaCentra](#) · [Aktuální stav modelů](#)

2 / Ruční sběr ve webchatu

Pro každou dvojici nový chat, dva snímky ve správném pořadí a společný P1. Vypněte web, nástroje, RAG, paměť a vlastní instrukce, pokud jsou dostupné. Celý finální text ukládejte podle postupu na str. 7.

3 / API pouze tehdy, pokud je používáte

Settings → Account → API keys → API key.

JWT nepoužívejte. Klíč uložte pouze lokálně a bezpečně, nikdy do reportu, CSV, screenshotu ani ZIPu.
`https://llm.ai.e-infra.cz/v1/`

API je volitelné; přiložené Python nástroje jsou offline, model samy nevolají. Podrobnosti: [oficiální dokumentace API](#).

Hromadný sběr na e-INFRA zahajte až po potvrzení vyučujícího. Nezávislé benchmarky vyžadují předchozí konzultaci s provozovatelem. Nevytvářejte paralelní dávky ani žebříček latence sdílené služby. [Pravidlo provozovatele](#).

Když se verze změní nebo zmizí

Zastavte dotčený sběr a informujte vyučujícího. **Starý a nový checkpoint neslučujte.** Zachovejte dosavadní výsledky; nový běh má oddělené označení a plán. Alias udržuje adresu služby, nikoli neměnné odpovědi.

O dostupnosti rozhoduje aktuální rozhraní; tento balíček neověřoval dostupnost přihlášeným API požadavkem. Klíč nesdílejte ani se spolužákem.

Pět dvojic a jeden prompt

Dvojice	Význam	Vyhodnocení
1 → 2, 1 → 3, 1 → 4	Tři nezávislá porovnání proti fotografii 1.	0/1/2 za každé změněné kritérium.
4 → 1	Obrácený směr; před/po se v klíči prohodí.	Stejná rubrika 0/1/2.
1 → 1	Dvakrát identický soubor fotografie 1.	Samostatná kontrola, bez bodů 0/1/2.

12 scén × 5 dvojic = 60 odpovědí na model.
Pět modelů: 300 odpovědí, z toho 240 změnových a 60 identických kontrol.

PROMPT P1 — kopírujte beze změn

Porovnej dvě fotografie stejné scény. První fotografie je stav PŘED a druhá fotografie je stav PO. Uveď všechny viditelné změny mezi nimi. U každé změny popiš stav před a po, včetně přesunu, otočení/orientace nebo zmizení/přidání objektu. Pokud něco nelze spolehlivě určit pouze z fotografií, napiš co a proč. Nevymýšlej změny, které z fotografií nelze ověřit.

V šabloně je přesný text v prompt.txt. Model nedostane klíč, počet změn, lidské otázky ani předchozí odpovědi. Nesmíte jej dodatečně navádět nebo žádat o opravu.

Která odpověď se započítá

První technicky úspěšná odpověď je jediná primární. Zachovejte celý finální text a všechny pokusy. Špatná, krátká, částečná odpověď, nejistota nebo odmítnutí jsou obsahové výsledky — neopakují se kvůli lepšímu skóre.

- Technická chyba:** timeout, nedostupná služba, síťová chyba, prokazatelně neodeslané snímky nebo chybějící finální text z technické příčiny. Zapište důvod, čas a pokus; žádné obsahové body.
- Nejvýše jeden technický retry na buňku.** Stejný model, vstupy a nastavení; nejdříve po 60 sekundách a po uplynutí delšího limitu služby. Úspěšnou odpověď už neopakujte.
- Trvalý výpadek:** po druhé chybě skončete; při potvrzené trvalé nedostupnosti i po první. Uložte failures.csv a vyžádejte rozhodnutí vyučujícího v exceptions.md. Bez schválení výjimky není nedokončené minimum splněno.

Plán zůstává nejméně 300 buněk. Počet pokusů může být vyšší, počet obsahových odpovědí při schváleném výpadku nižší. Výpadek není 0 bodů ani výsledek indeterminate. Cílené opakování úspěšných dotazů patří jen do odděleného rozšíření.

Jednotná pravidla vyhodnocení

A / Skóre skutečných změn

Body	Kdy je udělit za jedno kritérium
2	Správný konkrétní stav před i po; změna a její směr jsou určeny.
1	Správně rozpoznané, ale neúplné kritérium bez prokazatelně chybné části. Například jen správný stav po.
0	Chybné, chybějící nebo neurčité tvrzení bez identifikace změny. Správné „po“ nepřebije chybné „před“.

V `scores.csv` je jeden řádek na **skutečně změněné kritérium** dané dvojice a primární odpověď. Neměnné vlastnosti se nebudují. U 4-1 obraťte očekávané stavy. Ke každému skóre napište důvod.

B / Identická kontrola 1-1 — samostatně

control_result	Význam
correct_no_change	Model jednoznačně nehlásí relevantní změnu.
false_change	Tvrdí alespoň jednu konkrétní neexistující změnu; tato hodnota má přednost před neurčitou výhradou.
indeterminate	Nelze rozhodnout: odmítnutí, pouze nejistota nebo nejasný výsledek bez konkrétní tvrzené změny.

C / Ještě čtyři oddělené údaje na odpověď

- **Nepravdy navíc:** jeden samostatný ověřitelně chybný výrok = jedna nepravda. Opakování stejného výroku v téže odpovědi počítejte jednou. Chybu již potrestanou 0 u změnového kritéria znovu nepřičítejte. Neověřitelný detail není automaticky nepravda.
- **Nejistota:** `uncertainty_flag` = 1, pokud finální text výslovně připouští nejistotu nebo nemožnost určit detail; jinak 0. Nejde o počet slov „možná“.
- **Odmítnutí:** `refusal_flag` = 1 při výslovném odmítnutí úkolu, jinak 0. Zůstává platnou primární odpovědí; nepopsané změny mají 0, kontrola bývá indeterminate.
- **Technické chyby:** počet neúspěšných pokusů zvlášť; nedokončená buňka nemá obsahové skóre ani obsahové příznaky.

Souhrn = body / maximum dostupných změnových kritérií. Kontroly vykazujte ve třech kategoriích zvlášť. Uvedte pokrytí, počty nepravd, nejistot, odmítnutí a technických chyb. Při výpadcích srovnajte modely také na společném průniku z `common_summary.csv`; procenta z různých podmnožin nejsou spravedlivý přímý žebříček.

Od klíče po výsledek 3 / 4

Didaktický, smyšlený příklad. Není odpovědí žádného skutečného modelu ani novým měřením parapetu. Slouží pouze k nácviku zápisu a výpočtu; do vlastních výsledků jej nepřebírejte.

1 / Klíč předem: dvojice S01, 1 → 2

Kritérium	Před — foto 1	Po — foto 2	Boduje se?
C01 · poloha hrnku	Vlevo od knihy	Vpravo od knihy	Ano
C02 · špička pera	Nahoru	Dolů	Ano
C03 · barva knihy	Modrá	Modrá	Ne, stálost

2 / Celá ilustrativní odpověď

„Hrnek se přesunul zleva od knihy napravo. Pero má jinou orientaci, ale směr špičky nedokážu spolehlivě určit. Kniha změnila barvu z modré na zelenou.“

3 / Vysvětlení skóre

Kritérium	Body / maximum	Proč
C01	2 / 2	Poloha před i po správně a konkrétně.
C02	1 / 2	Rozpoznána změna orientace, chybí směry, bez chybné části.
C03	—	Nemění se; falešná změna barvy je jedna nepravda navíc.

Výsledek případu: 3 / 4 = 75 %.
extra_false_claim_count = 1 · uncertainty_flag = 1 · refusal_flag = 0 · control_result = NA · technical_retry_count = 0

4 / Jediné ID napříč soubory

Raw pokus v JSONL, řádky scores.csv a řádek response_metrics.csv mají totožné response_id: DEM0_M01_S01_1-2_A1. Koncovka A1 znamená první pokus. Nepoužívejte v jiném souboru zkrácené ID bez A1.

A jak by vypadala kontrola 1 → 1?

„Fotografie jsou stejné, žádné změny nevidím.“	correct_no_change
„Hrnek se přesunul doprava.“	false_change, 1 nepravda
„Nedokážu snímky porovnat.“	indeterminate

Strojově zapsaný příklad je v examples/. Jde o jeden ilustrační případ, ne hotové odevzdání s 300 odpověďmi. Výpočty celé práce vytváří offline nástroj ze všech vašich řádků.

Jak data získat a odevzdat

Rozbalte příloženou šablonu. Nástroje potřebují jen Python 3.10+ a standardní knihovnu. Podrobnosti a Windows/Linux varianty příkazů: **NAVOD.md**.

1 / Připravte a uzamkněte vstupy

Vyplňte project.json, modely, fotografie, klíč a validace. Vyučující přidělí jedinečné submission_id, aby se projekty nesmíchaly.

```
python validate_submission.py . --phase prepare
python workflow.py plan .
python workflow.py freeze .
```

2 / Každý skutečný pokus uložte

Nový chat, správný model, dvě fotografie a P1. Celý finální text zkopírujte do reply.txt. Pro každou buňku změňte kódy podle plan.csv. Nástroj nic neposílá AI.

```
python workflow.py record . --model M01 --scene S01 \
--pair 1-2 --response-file reply.txt
```

V terminálu můžete příkaz napsat na jediný řádek bez zpětného lomítka. Při technické chybě použijte --error-file error.txt. Pozdější přepis vyžaduje --timestamp se skutečným časem pokusu; žádné vymyšlené datum.

3 / Vyhodnoťte a přepočítejte

```
python workflow.py sheets .
# Nyní ručně vyplňte scores.csv a response_metrics.csv.
python workflow.py summary .
python validate_submission.py .
```

Prázdná pole nejsou automaticky nuly. Kontrola odmítne neúplná data, neplatné vazby, další pokus po úspěchu i nesprávné součty. Neověří význam odpovědi ani to, že člověk skutečně validoval snímky.

Co patří do ZIPu

Soubory / složky	Obsah
project.json, freeze.json, prompt.txt	Identita odevzdání, plán a uzamčené vstupy.
models.csv, ground_truth.csv, validation.csv	Modely, referenční klíč a dva spolužáci na scéně.
scenes/, responses/	Fotografie a celé raw JSONL všech pokusů.
scores.csv, response_metrics.csv, failures.csv	Ruční skóre, příznaky a případné technické výpadky.
summary.csv, common_summary.csv	Automatické souhrny včetně společného průniku.
README.md, report.pdf, extensions/E01/	Postup, zpráva a oddělené vlastní rozšíření.

Při výpadcích také exceptions.md s rozhodnutím vyučujícího. ZIP: ZUI2026_Prijmeni_Jmeno.zip. Hlavičky CSV neměňte; přesné datové typy a vazby jsou v schema.md. API klíče, .env a autentizační experty se nikdy neodevzdávají.

Zpráva a hodnocení studenta

Report: doporučeno 4–6 stran bez příloh

Cíl a dataset; modely a podmínky; metoda; souhrnná tabulka a graf/tabulka podle typů změn; nejméně tři konkrétní případy s ID, snímky, klíčem a citací; vlastní rozšíření; omezení a závěr. Použijte report_template.md.

Ukažte i úspěch a chybu, pokud obojí nastalo. Pokud ne, nevymýšlejte opačný případ. Celé odpovědi zůstávají v JSONL. Výsledky platí pro vaše scény a podmínky, ne pro všechny úlohy nebo všechny budoucí verze modelů.

Vlastní rozšíření — stačí smysluplná offline analýza

Jedna vlastní otázka nad rámec základní tabulky, popsání postup, reprodukovatelný výpočet či systematické porovnání, alespoň jedna tabulka/graf a závěr. Například porovnejte směry 1-4 a 4-1 u jednoho modelu ve všech 12 scénách.

Plán i výsledky patří do extensions/E01/. Rozlište předem plánovanou a explorativní otázku. Nový model, nový prompt či opakování jsou možné, ale nejsou povinné a nemíchají se do hlavní části.

35 bodů celkem · minimum 21 bodů

Max.	Oblast	Za co získáte plný počet
8	Fotografie a návrh scén	12+ vlastních, rozmanitých, čitelných scén; 4 stavy, stálý záběr, jasné změny a stabilní vlastnost.
7	Klíč a validace	Úplný klíč před AI; dvě doložené nezávislé kontroly každé scény; vyřešené nejasnosti.
8	Metoda a data	5+ různých checkpointů, P1 a jednotný postup, všechny pokusy, správné vazby a úplnost; žádný výběr lepších odpovědí.
7	Hodnocení a výsledky	Konzistentní 0/1/2, kontroly zvlášť, zdůvodnění, nepravdy/nejistoty/výpadky a správné souhrny.
5	Interpretace a rozšíření	Tři konkrétní případy, vlastní doložená analýza, omezení, srozumitelná zpráva a přiznané zdroje/pomoc.

V každé oblasti dostanete dílčí body podle splněných položek; přesné rozdělení je v HODNOCENI.md v šabloně a v metodice vyučujícího. Kreativita se hodnotí v návrhu scén, nikoli skrytým bonusem za „nachytání“ AI.

Chyba AI vám sama o sobě body nebere. Hodnotí se vaše příprava, poctivý postup a správná interpretace. Správně doložený a schválený výpadek se nehodnotí jako vaše obsahová chyba.

Podmínkou přijetí je dodržení hlavního rozsahu a pravidel, nebo výslovně schválená technická výjimka. Samotných 21 bodů nenahrazuje chybějící data či porušení integrity. Odevzdejte do **15. 11. 2026** podle pokynu vyučujícího.