



# Hitri seminar: Sodobna UI kot pomoč pri napornih pedagoških opravilih

Janez Perš · UL FE

*Za pedagoške kolege in asistente · Praktični delovni tokovi*



## Zakaj je to posebej pomembno za mlajše sodelavce?

**Mlajši sodelavci pogosto nosijo največji del operativnega pedagoškega bremena.**

Oddaje, vaje, napake, povratne informacije, zagovori in popraviljanje zahtevajo strokovno presojo, a se pogosto ponavljajo.

Današnja teza: UI je koristna tam, kjer ohrani odgovornost pri učitelju, zmanjša ponavljanje in poveča preglednost dela.

# Ocenjevanje ni drobna administracija



*Je strokovno delo, ki se ponavlja, se kopiči in tiho poje čas za boljše poučevanje.*

# 9,9

**ur / teden**

Povprečen čas, ki ga učitelji v anketi porabijo za ocenjevanje in označevanje.

--

**energija**

Obsežno ocenjevanje porablja pozornost, ki bi lahko šla v načrtovanje predmeta.

!

**čustveni strošek**

Na univerzi je ocenjevanje doživeto manj pozitivno kot raziskovanje in poučevanje.

Learnosity (2025): učitelji poročajo o 9,9 h ocenjevanja na teden (n=258).  
Schwab et al. (2024): ocenjevanje je čustveno manj pozitivno kot raziskovanje in poučevanje.  
Terada in Merrill (2024): pogosto ocenjevanje jemlje čas in energijo za izboljševanje pouka.

# V STEM predmetih je breme še težje



*Ne preverjamo samo končnega odgovora, ampak pot, skice, formule, napake in razloge.*

## Več korakov

Študent lahko pride do napačnega rezultata z dobro metodo ali do prave številke po slabi poti.

## Več modalnosti

Rokopis, diagrami, vezja, enačbe in prosti komentarji so del odgovora, ne okras.

## Več zakasnitve

Ko se povratna informacija zavleče, je manj uporabna za učenje in naslednje vaje.

Tan et al. (2025): STEM ocenjevanje vključuje večkorakovne rešitve, formule, diagrame in grafične artefakte.

Tan et al. (2025): pri večjih obremenitvah je pravočasna povratna informacija poseben problem.

Singh et al. (2017): digitalizacija izboljša hitrost in konsistentnost, presoja pa ostane človeško delo.

# Teza seminarja: razbremeniti ponavljanje, ne presoje



UI je uporabna šele, ko je delo ponovljivo, preverljivo in popravljivo.

Priprava	Preverjanje	Ocenjevanje	Dokumentiranje
<b>iz gradiv</b>	<b>oddaj</b>	<b>STEM odgovorov</b>	<b>odločitev</b>
<b>Hitrejši prvi osnutek</b>	<b>Boljša vprašanja</b>	<b>Več nadzora</b>	<b>Več sledljivosti</b>
Viri, slike, struktura in poudarki se uredijo v pregledno izhodišče.	Sistem pripravi izhodišča za pogovor, učitelj pa vodi presojo.	Reference, pravila, več ocenjevalcev in ročni pregled zmanjšajo slepe odločitve.	Vsaka trditev mora imeti vir, razlog in možnost popravka.

Singh et al. (2017): urejen digitalni delovni tok zmanjša režijo in izboljša konsistentnost.

Tan et al. (2025): avtomatizacija je smiselna tam, kjer zmanjša obremenitev in pospeši povratno informacijo.

Schwab et al. (2024): zmanjševati je treba tudi čustveni strošek ocenjevanja, ne samo število klikov.

# Meja je jasna: orodje ne sme prevzeti odgovornosti



*Najslabša uporaba UI je tista, ki pospeši delo tako, da skrije presojo.*

**Ne**

UI ne pozna lokalnih pravil, ciljev predmeta in mejnih primerov sama od sebe.

**Ne**

Modelov odgovor ni odločba; je gradivo za presojo, preverjanje in izboljšavo.

**Ne**

Detektorji niso stabilna pedagoška podlaga za obtoževanje študentov.

**Ne**

Če postopka ni mogoče razložiti študentu, ni zrel za pedagoško rabo.

Singh et al. (2017): uspešni sistemi najprej uredijo človeški delovni tok.

Tan et al. (2025): avtomatizacija naj zmanjša obremenitev in pospeši povratno informacijo, ne odstrani odgovornosti.

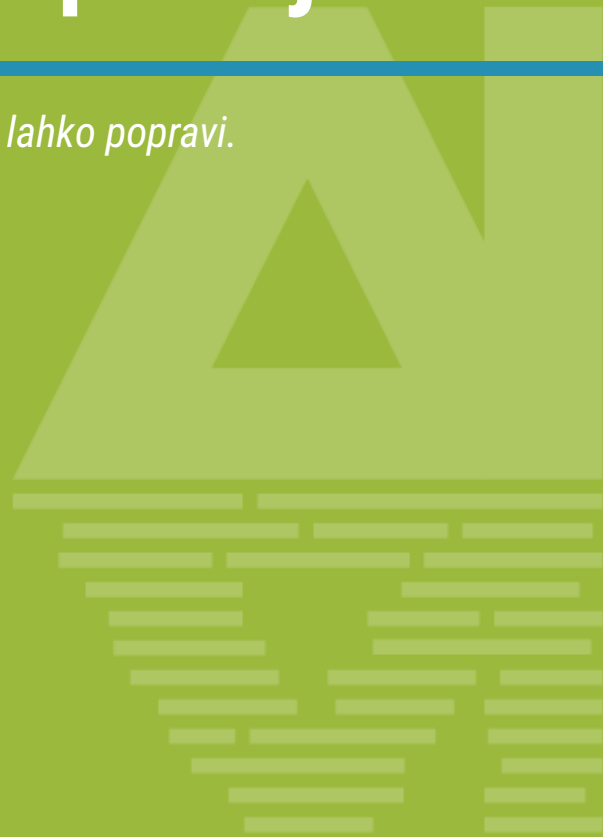
Schwab et al. (2024): zmanjšanje bremena ne sme razvrednotiti strokovne presoje.

Primer 1

# UI-podprta priprava prosojnic za STEM

---

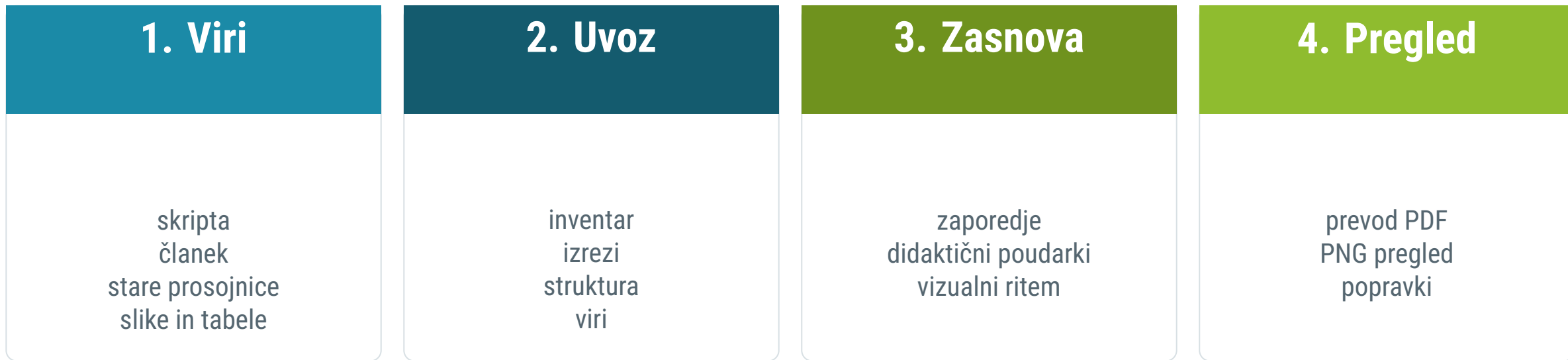
*Od gradiv do preverljivega osnutka, ki ga učitelj lahko popravi.*



# Postopek priprave prosojnic iz poljubnih gradiv



Repozitorij vsebuje navodila agentu ter spretnosti za uvoz virov, izdelavo predstavitev in razvoj predlog LaTeX.



- GitHub repozitorij: <https://github.com/ulfe-lmi/slaif-presentation-builder>
- Agent iterira izgled, dokler izgled ni v redu.
- Učitelj ostane avtor: določi raven razlage, izbriše odvečno, popravi poudarke in potrdi končno sporočilo.
- Učitelj komunicira z agentom; ni treba, da se sploh zaveda, da je v ozadju LaTeX.

# Uvoz vira: kaj mora agent oddati



*Dober uvoz ni povzetek, ampak mapa dokazov za kasnejšo predstavitev.*

# 1

## Struktura

kaj je v virih in kako bi to lahko postalo zaporedje prosojnic

# 2

## Gradniki

slike, izrezi, tabele, strani in drugi elementi, ki jih je smiselno uporabiti

# 3

## Sledljivost

od kod prihaja trditev ali vizualni element, da učitelj lahko preveri vir

- To ni samo formalnost.
- Dobre prosojnice so kompromis med strukturo in izgledom.
- Rezultat "ChatGPT, naredi mi PPTX" je lep, ampak nekonsistenten izdelek - popolnoma brez strukture.
- Struktura koristi, ko hočemo gradivo ponovno uporabiti ali popraviti.

# Ustvarjanje predstavitev: kaj mora agent preveriti



*Uspešen prevod LaTeX-a še ne pomeni uporabne predstavitve.*

## Vsebina

slovenščina  
jasno zaporedje  
primerna gostota

## Vizualno

berljive slike  
pravilni izrezi  
dovolj praznega prostora

## Prevod

lokalni PDF  
brez napak  
ponovljiv ukaz

## Pregled

PNG izris  
ročno gledanje  
popravki

### Končni standard

*Agent mora sam iskati prekrivanje besedila, premajhne črke, slabe izreze in zlomljene postavitve.*

## Primer 2

# Priprava vprašanj za zagovor samostojno izvedenih aktivnosti

---

*Kako preverimo, ali študenti razumejo oddane izdelke.*

# Študenti uporabljajo umetno inteligenco



Zanesljive detekcije uporabe UI ni, zato mora zagovor preverjati razumevanje konkretne oddaje.

## Slaba smer

"Ali si uporabil UI?"

Nezanesljivo, težko dokazljivo in pogosto pedagoško neproduktivno.

## Dobra smer

"Ali razumeš to, kar si oddal?"

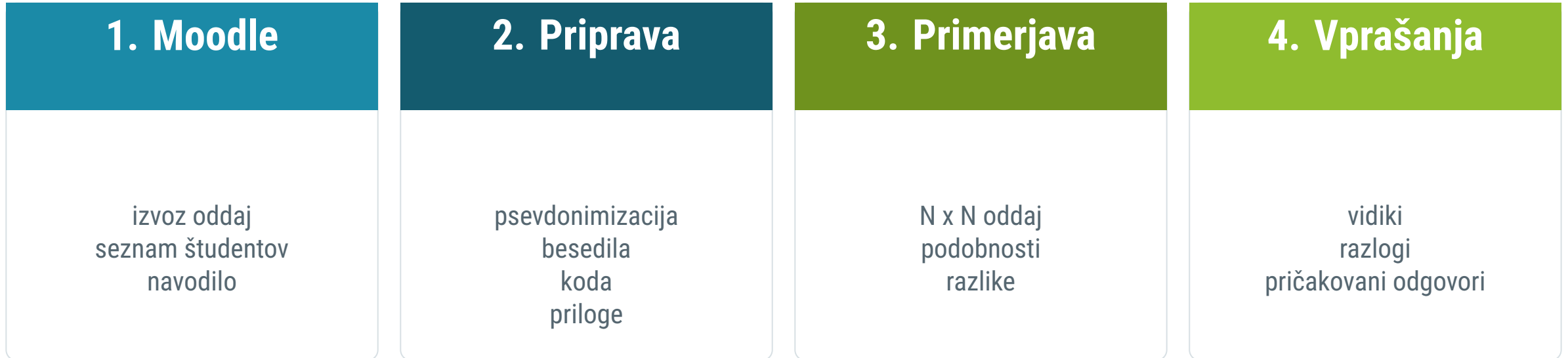
Vprašanja o kodi, grafu, parametru, odločitvi, napaki ali nenavadni razlagi.

- GitHub repozitorij: <https://github.com/ulfe-lmi/slaif-defense-grading>
- Ne dokazujemo uporabe UI, ampak preverjamo razumevanje oddanega izdelka.
- Slab zagovor je pedagoško pomemben signal, ne glede na vzrok.
- UI pripravi ciljana vprašanja, ne sklepa o avtorstvu.

# Delovni tok za zagovore vaj



Oddaje se izvozijo iz Moodlea, psevdonimizirajo, primerjajo  $N \times N$  in pretvorijo v vprašanja za zagovor.



- Agent sistematično izvede  $N \times N$  primerjav oddanih izdelkov.
- Izhod niso sodbe, ampak podlaga za ciljana vprašanja.
- Vsako vprašanje vsebuje razlog, vprašanje študentu in pričakovani odgovor.

# Kakšno vprašanje je uporabno na zagovoru?



*Dobro vprašanje je specifično za študenta, preverljivo in vezano na njegovo oddajo.*

## Preširoko

Razloži regresijo.

Ali razumeš program?

Kaj je rezultat?

## Uporabno

Zakaj si tukaj izbral ta prag?

Kaj se zgodi, če zamenjamo parameter?

Zakaj je ta graf drugačen od pričakovanega?

- Vprašanja so specifična za posameznega študenta in njegovo konkretno oddajo.
- Kjer je mogoče, se vprašanja ne ponavljajo.
- Standardni naučeni odgovori izgubijo vrednost.
- Pedagog dobi individualna vprašanja brez ročne priprave za vsakega študenta.

## Primer 3

# UI-podprto ocenjevanje STEM izpitov

---

*Do GPT-5.2 je bil to doslej nerešen problem.*



# Ročno pisani izpiti v tehniki in naravoslovju



Ocenjevanje mora biti iz perspektive študenta 100% kompatibilno z dosedanjim načinom pisnih izpitov.

## Vhod je kompleksen

ročno pisanje  
enačbe  
diagrami  
popravki

## Ocena je postopna

vmesni koraki  
točke po pravilih  
razlaga napak

## OCR ni dovolj

matematika  
skice  
nečitljivost  
kontekst

## Zato: sistem

referenca  
pravila  
več ocen  
nadzor

### Zakaj je to dober primer

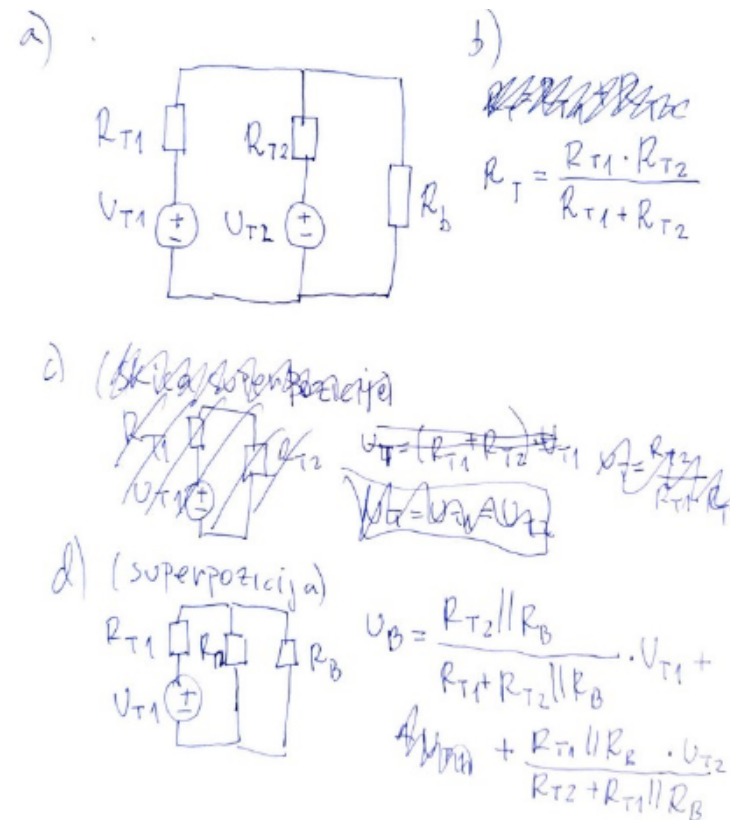
*Pri ročno pisanih izpiti "samo uporabi model" hitro odpove; urejen cevovod pa lahko pomaga, ker loči prepoznavanje, pravila, ocenjevanje, validacijo in ročni pregled.*

# Zakaj je STEM rokopis drugačen problem



Sken pisnega izpita je zelo kompleksen ročno narejen dokument.

- besedilo in formule sta prepletene;
- diagram je del odgovora;
- nekaj je prečrtano ali popravljeno;
- ocena mora razložiti delne točke.

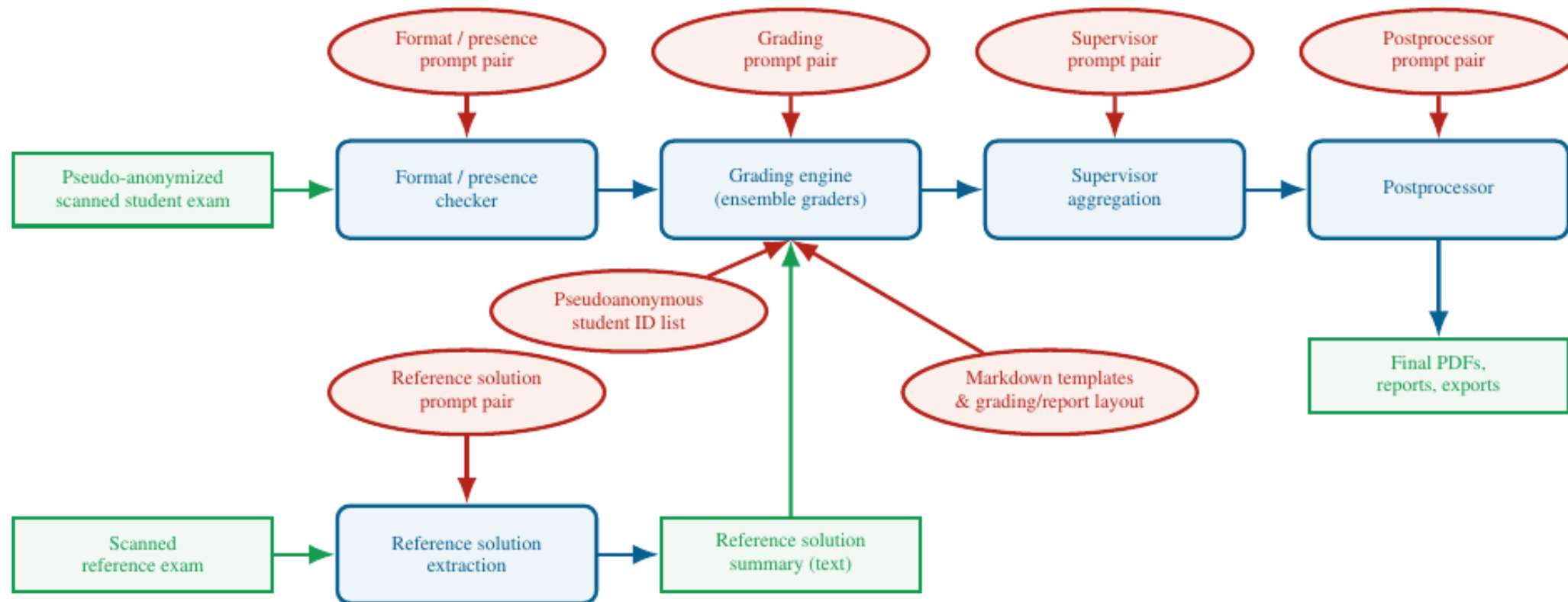


Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

# Arhitektura: ne en prompt, ampak cevovod



GitHub repo še ni javen, na željo lahko dobijo dostop zainteresirani pedagogi.



Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

# Kaj je novost in zakaj je evalvacija težka



*Pomemben rezultat ni "model zna ocenjevati", ampak sistemska zasnova okoli modela.*

## Novost

realistični skeni STEM rešitev  
brez prilagajanja študentov  
brez ročnega pretipkavanja

## Sistemska zasnova

referenčna rešitev  
pravila in stroge predloge  
več neodvisnih ocen  
nadzorni korak

## Težka evalvacija

zasebnost podatkov  
redki javni nabori  
človeška ocena ni absolutna  
pomembna je tudi razlaga

- Poleg točnosti je pomembna tudi nepristranskost ocenjevanja.
- Rezultati morajo biti stabilni med ponovitvami.
- Sistem mora zaznati primere, ki zahtevajo ročni pregled.
- Povratna informacija mora študentu razložiti, zakaj je rešitev pravilna, delno pravilna ali napačna.

# Referenčna rešitev: zakaj je ključna



*Ocenjevanje ni "model proti študentu", ampak model ob pravilih in lokalni normi.*

## 1. Referenca

Predavatelj pripravi 100% rešitev, lahko tudi ročno pisano.

## 2. Povzetek

Sistem pretvori referenco v ocenjevalno sidro.

## 3. Ocenjevanje

Odgovor se presoja ob pravilih, ne ob splošnem vtisu modela.

### **Zakaj referenca ni dodatek**

*Brez reference se model lažje premakne v splošno razlago in sistematično precenjevanje.*

# Prisotnost odgovora in stroge predloge



*Dve majhni varovali, ki preprečita velik del nereda.*

## Prisotnost odgovora

Najprej preveri, ali ima vprašanje dejanski odgovor. Praznine se ne smejo spremeniti v delne točke.

## Stroga predloga

Poročilo ima pričakovano strukturo: vprašanje, povzetek, pravila, razlaga in točke.

## Validacija izvoza

Če vsota ne ustreza prispevkom ali format ni strojno berljiv, sledi ročni pregled.

### Zakaj stroga oblika šteje

*Cevovod zaradi svoje zasnove izdela strukturiran zapis, ki ga lahko deterministično razčlenimo, preverimo, preračunamo in prenesemo v tabele ali poročila.*

# Zakaj več ocenjevalcev in nadzornik?



*En sam modelni odgovor je premalo stabilen za pedagoško rabo.*

Ocenjevalec A	Ocenjevalec B	Ocenjevalec C	Nadzornik
neodvisen osnutek	neodvisen osnutek	neodvisen osnutek	združi osnutke, uveljavi predlogo in označi neskladja

## Interpretacija

*Nesoglasje ni napaka; je signal za človeški pregled.*

# Primer referenčne rešitve



Referenčna rešitev model zasidra na pravo mesto glede pričakovanega nivoja znanja in podrobnosti.

- ni nujno tipkana; lahko je ročno pisana;
- vsebuje metode, skice in formule;
- služi kot lokalna ocenjevalna norma;
- mora biti dovolj jasna, da jo sistem povzame.

⇒ Ime in Priimek:  
ROŠTJAN MUROVEC

Kolokvij EDT – teden 2

1. (25 %) Pri uporabi napetostnih delilnikov smo naleteli na kompromis: iz določenih razlogov želimo delilnik zgraditi iz uporov čim manjših upornosti, po drugi strani pa želimo, da so upornosti čim večje. Razložite protislovje in razloge zanj.

- THEVENINOVA UPORNOST DELILNIKA JE  $R_1 || R_2$ , KJER STA  $R_1$  IN  $R_2$  UPORA, KI GA SESTAVLJATA. ŽELIMO, DA JE THEVENINOVA UPORNOST ČIM MANJŠA, DA ŽEMANJŠAMO SESEJANJE DELILNIKOVE NAPETOSTI,  $\Rightarrow$  ČIM MANJŠI UPORI.
- DELILNIK POABLAJA TOK VIRA  $V_1$ , KI JE ENAK  $V_1 / (R_1 || R_2)$ . MANJŠI KOT SO UPORI VEČ TOKA POABLAJA IN USTVARJA VEČJE IZSEBE.  $\Rightarrow$  ČIM VEČJI UPORI.

2. (25 %) kateri pogoj mora biti izpolnjen ob priklopu bremena na napetostni delilnik iz uporov  $R_1$  in  $R_2$ , da je breme krmiljeno tokovno?

BREME JE KRMILJENO TOKOVNO, KO JE BREMENSKA UPORNOST  $R_B$  MANJŠA OD THEVENINOVE UPORNOSTI DELILNIKA  $R_T$  ( $R_B \ll R_T$ ).

$$R_T = R_1 || R_2 \Rightarrow R_B \ll (R_1 || R_2)$$

Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

# Izhod ni samo številka, ampak obsežen dokument



Glavna prednost za študente: obsežna povratna informacija s podrobno razlago, kaj so naredili narobe.

## Product Evaluation

# 45%

**Final grade:** 45%

**First name Last name:** James Bond

**Student ID number:** 100007

**Course:** CVWW

**Exam:** CVWWTest

**Evaluation date:** 23.12.2025

**Partial grades:** 50%, 40%, 47.5%

**Evaluator model:** model=gpt-5.2, generator\_effort=high, supervisor\_effort=high, generator\_temperature=0.55, supervisor\_temperature=0.1, ensemble\_size=3

**Software version:** batch\_ensemble\_grade=commit e8f4b43

(e8f4b436b84f5bf1c0a8aa56639a50d5019072fa) +dirty, try\_deanonymize=commit e8f4b43

(e8f4b436b84f5bf1c0a8aa56639a50d5019072fa) +dirty,

results\_to\_excel=commit e8f4b43 (e8f4b436b84f5bf1c0a8aa56639a50d5019072fa)

+dirty, export\_final\_pdfs=commit e8f4b43

(e8f4b436b84f5bf1c0a8aa56639a50d5019072fa) +dirty

The list of answered questions includes only Questions 1 and 2, while the graders' reports also include Task 3 (split into 4 subtasks). Since no rows are provided for the additional questions in the list, only tasks with known weights and contributions from the reports are taken into account in the grading.

## Task 1

### Question

**1. (25%) When using voltage dividers, we encountered a trade-off: for certain reasons we want to build the divider from resistors with as small resistance as possible, but on the other hand we want the resistances to be as large as possible. Explain the contradiction and the reasons for it.**

### Student's answer

Describes the trade-off: small resistors "because of voltage", large resistors "because of current".

### Correctness explanation

The answer does indicate the existence of a trade-off, but it does not demonstrate understanding by explaining the key reasons (e.g., the Thevenin resistance of the divider ( $R_T = R_1 \parallel R_2$ ), the effect of loading on the "collapse" of the output voltage, and the divider current ( $I = U_1 / (R_1 + R_2)$ ) and the resulting losses/consumption). Therefore, the solution is too general and only partially justified.

References: [R1, R6]; EXAMPLE: NO; SKETCH: —; CROSSED OUT: NO

### Scoring

20% of 25% → contribution: 5.0%

## Task 2

### Question

Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

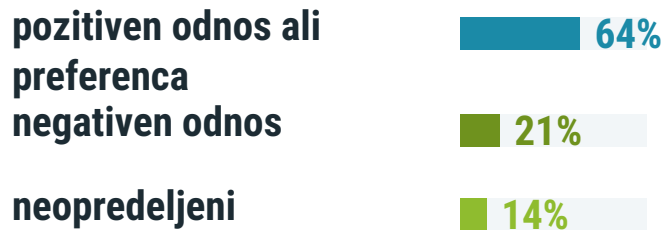
# Povratna informacija od študentov



Predmet: Elektronika z digitalno tehniko, zimski semester 2025/2026

## Vprašanje 1

Odnos do UI-prvega ocenjevanja



## Vprašanje 2

Ali imate korist od sistema?



## Vprašanje 3

Katere koristi vidite?



Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026. V anketi je sodelovalo 14 študentov po 8 tedenskih UI-ocenjenih kvizih.

# Povratna informacija od študentov



Predmet: Elektronika z digitalno tehniko, zimski semester 2025/2026

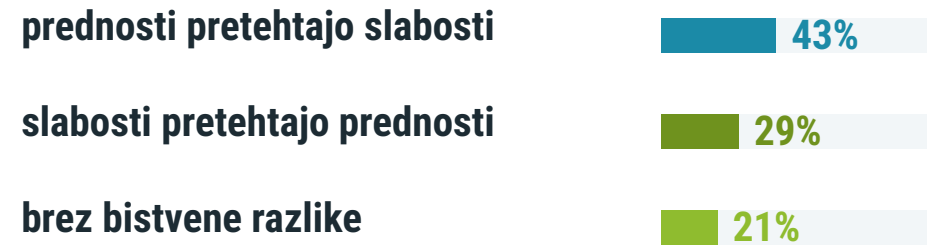
## Vprašanje 4

### Kaj vas skrbi?



## Vprašanje 5

### Skupna presoja



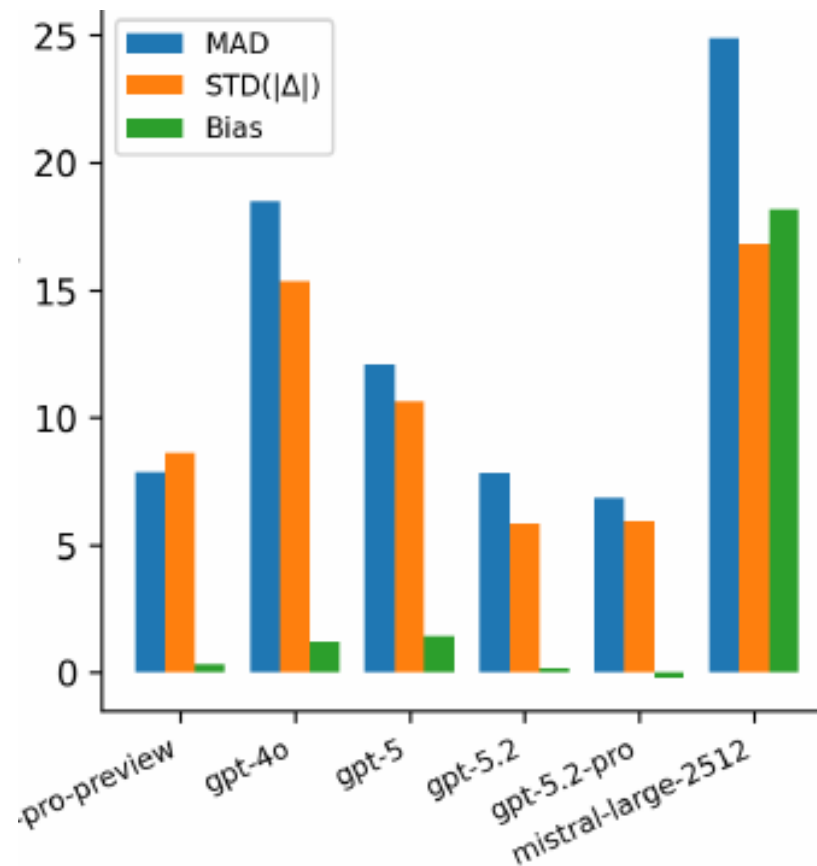
Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026. V anketi je sodelovalo 14 študentov po 8 tedenskih UI-ocenjenih kvizih.

# Kaj povedo rezultati



Najboljši sodobni modeli so uporabni, šibkejši pa hitro odpovejo.

- MAD: povprečna absolutna razlika do učiteljeve ocene;
- $STD(|\Delta|)$ : razpršenost absolutnih napak med študenti;
- bias: sistematično precenjevanje ali podcenjevanje;
- najboljši sistemi dosežejo enomestni MAD z nizkim biasom.



Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

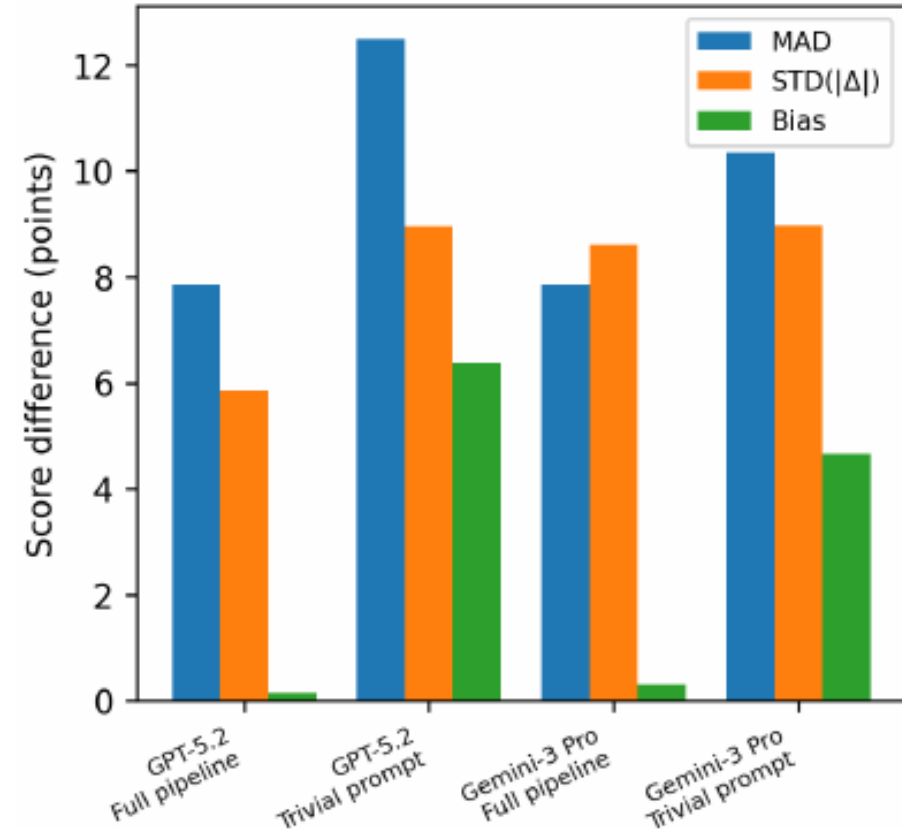
# Zakaj "samo vprašaj model" ni dovolj



Ablacija: trivialno pozivanje brez pravil, reference in nadzornika poslabša ocenjevanje.

- referenca ni dodatek, ampak sidro;
- pravila morajo biti vidna v razlagi;
- predloga prepreči formatni kaos;
- nadzornik in validacija zmanjšata zdrse.

Effect of pipeline guidance vs trivial prompting



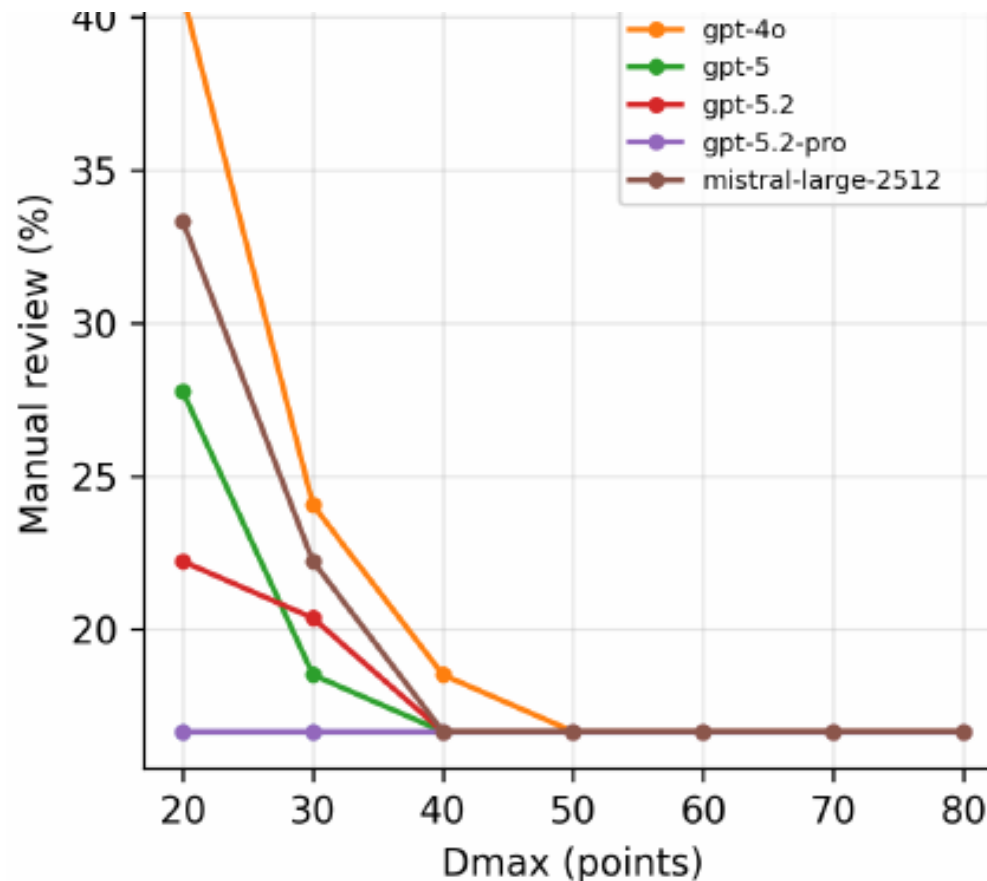
Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

# Ročni pregled kot varnostni ventil



Ne avtomatiziramo vsega; avtomatiziramo pripravo in označimo sporne primere.

- kadar se ocenjevalci ne strinjajo, naj človek pogleda;
- prag ni tehnična malenkost, ampak pedagoška odločitev;
- cilj je zmanjšati obseg pregleda, ne odstraniti odgovornost;
- tako dobimo nadzorovan kompromis med hitrostjo in varnostjo.



Perš, J., Muhovič, J., Košir, A., & Murovec, B. (2026). *Grading Handwritten Engineering Exams with Multimodal Large Language Models*. arXiv:2601.00730v1 [cs.CV], 2. januar 2026.

# Etika in lokalna pravila



*Sistem mora biti tak, da ga predavatelj razume dovolj dobro, da ga lahko zavrne.*

## Podatki

psevdoanonimizacija, lokalna deanonimizacija in jasen nadzor nad vhodnimi datotekami

## Študent

ve, kateri del postopka je avtomatiziran in kje odloča človek

## Napake

sistem pričakuje napačno branje, spregledan diagram in preveč samozavestno razlago

### Minimalni standard

*V izobraževanju ni dovolj, da je ocena približno pravilna; postopek mora biti razložljiv in popravljiv.*

# Kaj bi bil dober naslednji korak pri nas?



*Ne velika obljuba, ampak majhen, preverljiv pilot.*

## 1. Izberemo

en predmet, eno vrsto naloge in delovni tok, ki res obremenjuje ekipo

## 2. Zberemo

navodila, primere oddaj, kriterije in stare povratne informacije

## 3. Naredimo

osnutek prosojnic, vprašanj za zagovor ali ocenjevalnega poročila

## 4. Merimo

čas  
napake  
popravljenost  
nadzor

### Naslednji korak

*Cilj jesenske delavnice: konkretna opravila iz vaših predmetov, ne abstraktna razprava o UI.*

*Hvala za pozornost!*



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE

SLAIF: Slovenska tovarna umetne inteligence je prejela financiranje Skupnega podjetja za evropsko visokozmogljivo računalništvo v okviru sporazuma o nepovratnih sredstvih št. 101254461. Projekt financirata Ministrstvo za visoko šolstvo, znanost in inovacije Republike Slovenije ter program Horizon Europe Evropske unije.

SLAIF: Slovenian AI Factory has received funding by the European High Performance Computing Joint Undertaking under the grant agreement no. 101254461. The project is funded by the Ministry of Higher Education, Science and Innovation of Republic of Slovenia and by Horizon Europe Programme of the European Union.