

## Образац 1 – Пријава теме докторске дисертације

(Кандидат доставља пријаву Програмском савету)

Назив студијског програма: Интелигентни системи

Кандидат: Марко Путниковић

Молим да ми се одобри израда докторске дисертације под насловом:

“Мултиагентски приступи засновани на великим језичким моделима за објашњиво аутоматско оцењивање студентских радова”

### 1. Биографија кандидата

Марко Путниковић рођен је 13.5.1993. године у Ваљеву, Србија. Уписао је основне студије на Факултету организационих наука Универзитета у Београду 2013. године, смер Информациони системи и технологије. Дипломирао је 2017. године стекавши звање Инжењер организационих наука - информациони системи и технологије. Мастер студије је уписао 2017. године на Универзитету у Греноблу, на смеру Вештачка интелигенција и семантички веб, као стипендиста владе Француске. Мастер студије је завршио 2018. године стекавши звање Мастер информатичар. Током мастер студија је обавио петомесечну праксу (stage) при Лабораторији за информатику Универзитета у Греноблу. Докторске студије је уписао 2019. године на Универзитету у Београду на програму мултидисциплираних студија при универзитету, смер Интелигентни системи. Током студија се бавио применом метода процесирања природних језика у области едукације и објавио истраживачки рад у M21 часопису о примени тих метода у оцењивању студентских радова (2.1.1, 2.2.1).

Упоредо са докторским студијама, кандидат је радио у компанији NCR као софтвер инжењер 2019-2022. године, након чега прелази у компанију Robert Bosch у којој и даље ради као инжењер вештачке интелигенције. У компанији Robert Bosch је радио на иновационом истраживачком пројекту о примени метода процесирања природних језика за екстракцију података са веба, у оквиру ког је објавио истраживачки рад на конференцији (2.2.2).

### 2. Библиографија кандидата (категорисано према категоризацији надлежног Министарства)

#### 2.1. Објављени радови или прихваћени за штампу (DOI број или изјава уредника)

Истакнути међународни часописи (M21):

- 2.1.1. Putnikovic, M., & Jovanovic, J. (2023). Embeddings for automatic short answer grading: A scoping review. *IEEE Transactions on Learning Technologies*, 16(2), 219-231.

#### 2.2. Конференције и скупови

- 2.2.1. Putnikovic, M., & Jovanovic, J. (2022). *Word Embeddings for Automatic Short Answer Grading* (pp. 197-213). World Scientific Publishing Co. Pte. Ltd.
- 2.2.2. Hätyy, A., Milchevski, D., Döring, K., Putnikovic, M., Mesgar, M., Novović, F., ... & Stranjanac, I. (2024, November). A Cost-Efficient Modular Sieve for Extracting Product Information from Company Websites. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track* (pp. 1444-1456).
- 2.2.3. Stanković, R., Rađenović, J., Škorić, M., & Putnikovic, M. (2025, March 27). Learning Word Embeddings using Lexical Resources and Corpora. ICIST - International Conference on Information Society and Technology (ICIST 2025), Kopaonik, Serbia. Zenodo. <https://doi.org/10.5281/zenodo.15093900>

### 3. Предмет и циљеви докторске дисертације

#### 3.1 Предмет докторске дисертације (максимално 1 страна)

Предмет докторске дисертације је аутоматско оцењивање студентских радова и пружање објашњења о процесу оцењивања применом мултиагентских система заснованих на великим језичким моделима.

Оцењивање студентских радова представља један од главних процеса у оквиру образовног система, јер омогућује процену и пружа повратну информацију о усвојености знања студента. Оцењивање може бити сумативне природе, што подразумева доделу оцене и даје формалну основу за напредовање и стицање академског звања; и формативне природе, које укључује пружање повратне информације студенту, праћење напретка у учењу и усмеравање даљег развоја компетенција студента. За наставника, задатак оцењивања је обично временски ограничен и когнитивно напоран у зависности од броја студената, што може довести до замора, утицати на доследност и смањити могућност боље подршке студентима. Додатно, субјективност наставника, тачније, лична уверења и очекивања, ослањање на претходно искуство и интуицију може довести до неконзистентности при оцењивању и утицати на исход. Са развојем онлајн облика образовања, све већи значај добијају системи за аутоматско оцењивање, чија је намена да се омогући брже и конзистентније оцењивање великог броја студентских радова, као и да наставнике растерети дела процеса оцењивања и пружања образложења додељене оцене студенту који се може аутоматизовати.

Формално, задаци који се оцењују се могу поделити на оне затвореног типа (нпр. задатак са више понуђених одговора) и отвореног типа (конструисан одговор) (Bennet, 1991). Задаци затвореног типа подразумевају одабир једног или више понуђених одговора, при чему је оцењивање детерминистичко, тј. засновано на унапред дефинисаном решењу. С друге стране, задаци отвореног типа се могу поделити на оне са кратким и дугим текстом слободног одговора (есеји), али и у зависности од домена предмета могу бити програмски код, математички доказ и сл. За отворен тип одговора, који ће бити фокус истраживања ове докторске дисертације, потребна је обрада и анализа садржаја, што захтева примену софистициранијих метода процесирања природних језика.

Историјски, методе оцењивања задатака отвореног типа се могу поделити на оне (1) засноване на шаблонима и правилима, (2) статистичким карактеристикама текста, и (3) машинском и дубоком учењу (Burrows et al. 2015). Приступ заснован на дубоком учењу је посебно присутан последњих година услед значајног успеха примене великих језичких модела на оцењивање студентских радова (Fagbohun et al. 2024). Међутим, у великом броју постојећих истраживања све задатке у оквиру процеса оцењивања обавља један велики језички модел, што може довести до разних проблема који се генерално приписују таквим моделима, као што су халуцинације и губитак контекста. Осим тога, оцењивање једним великим језичким моделом може довести до мешања критеријума за оцењивање уколико је потребно оценити различите аспекте рада понаособ. Додатно, информације које један велики језички модел поседује могу бити застареле или нерелевантне, те је потребно обогатити упит текућим знањем битним за контекст оцењивања.

Приступ којим би се могао унапредити процес оцењивања и пружања објашњења тог процеса помоћу великих језичких модела јесте примена система оцењивања заснованих на више језичких модела у међусобној спрези. Такви системи се називају мултиагентским системима (МАС), где је агент софтверска компонента која, између осталог, користи велики језички модел како би обавила одређени задатак у процесу оцењивања (Kamalov et al. 2025). Поред великих језичких модела, агент може користити разне друге програмске алате како би обогатио своје текуће знање о релевантном контексту и дао бољи резултат (претрага веба, библиотеке за анализу података, приступ базама података и сл.). Примена МАС-а заснованих на великим језичким моделима у аутоматском оцењивању и даље представља релативно нову и недовољно систематски истражену област. Посебно недостају истраживања која анализирају утицај архитектуре, тј. структурне организације МАС-а на резултате оцењивања. Конкретно, ограничен је број истраживања која испитују предности и недостатке мултиагентских приступа у односу на друге системе аутоматског оцењивања, испитују утицај специјализације агената по различитим аспектима оцењивања на тачност оцењивања, начине агрегације оцена појединачних агената у коначну оцену, и анализирају ефекте интеграције спољног знања у МАС. Поред тога, постојећа истраживања се недовољно баве питањем транспарентности и објашњивости процеса оцењивања у МАС-у. Образложење оцене коју додељује аутоматски систем неопходно је како би се обезбедила транспарентност оцењивања и омогућила провера примене рубрика за оцењивање, чиме се гради поверење у систем. Додатно, регулаторни правни оквири попут Акта о вештачкој интелигенцији Европске уније (European Union, 2024, Article 86) захтевају одређени ниво транспарентности аутоматског одлучивања, посебно у доменима примене са високим ризиком у које се убраја и образовање.

### **3.2. Циљеви докторске дисертације (максимално 1 страна)**

Општи циљ ове докторске дисертације је развој тачног, конзистентног и објашњивог МАС-а заснованог на великим језичким моделима за аутоматско оцењивање студентских радова. У складу са дефинисаним општим циљем, формулисани су следећи конкретни циљеви:

1. Испитати у којој мери специјализација агената у односу на рубрику доприноси тачности аутоматског оцењивања
2. Испитати допринос интеграције релевантног спољног знања (нпр. курикулума или наставних материјала) на тачност аутоматског оцењивања

3. Анализирати како различiti начини агрегације процена појединачних агената утичу на тачност аутоматског оцењивања
4. Обезбедити транспарентност и објашњивост процеса оцењивања МАС-а кроз генерисање утемељених и аргументованих објашњења

### 3.3 Хипотезе

На основу анализе релевантне и расположиве литературе, као и уочених недовољно истражених тема у области аутоматског оцењивања, формулишемо следеће хипотезе:

- X1. Примена МАС-а у аутоматском оцењивању студентских радова доводи до повећања тачности оцењивања у односу на постојеће приступе.
- X2. Адаптивна специјализација агената у односу на тип задатка, заснована на рубрици за оцењивање, доводи до повећања тачности аутоматског оцењивања у односу на мултиагентски приступ без адаптивне специјализације агената.
- X3. Интеграција релевантног спољног знања у МАС, одређена типом задатка који је предмет оцењивања, доводи до веће тачности аутоматског оцењивања, у односу на МАС без употребе спољног знања.
- X4. Увођење агената за агрегацију у МАС, који на основу процена појединачних агената доноси коначну оцену, доводи до веће тачности аутоматског оцењивања у односу на МАС у којем се коначна оцена добија једноставном агрегацијом (нпр. упросечавањем) оцена појединачних агената.
- X5. Мултиагентски систем заснован на великим језичким моделима омогућује већу транспарентност и објашњивост процеса аутоматског оцењивања у односу на моноагентски приступ, кроз генерисање образложења донете оцене.

### 4. План рада

План рада на докторској дисертацији ће се састојати из следећих корака:

- преглед релевантне литературе из области аутоматског оцењивања, великих језичких модела и МАС-а
- анализа постојећих примена МАС-а заснованих на великим језичким моделима за аутоматско оцењивање
- избор јавно доступних скупова података за евалуацију аутоматског оцењивања есеја и кратких одговора, као и припрема скупа података за оцењивање програмског кода
- Прикупљање рубрика за оцењивање по типу задатка прегледом релевантне литературе и анализом рубрика из јавно доступних скупова података
- дизајн моноагентског система за аутоматско оцењивање (базни модел)
- дизајн МАС-а са адаптивном специјализацијом агената
- дизајн МАС-а са агентом за агрегацију оцена
- дизајн МАС-а са приступом спољном знању
- укључивање механизма за генерисање утемељених и аргументованих образложења оцене у оквиру МАС-а ради обезбеђивања транспарентност и објашњивости процеса оцењивања
- одабир репрезентативних великих језичких модела који ће се користити у експерименталној евалуацији
- дефинисање метрика за евалуацију, тј. процену тачности дизајнираних МАС-а, као и процену квалитета аутоматски генерисаних образложења оцена
- имплементација предложених система и експерименталних поставки
- извођење експеримената и прикупљање резултата
- анализа резултата, визуелизација и интерпретација у односу на постављене хипотезе
- дискусија ограничења истраживања и предлог будућих праваца развоја

### 5. Библиографски подаци релевантни за докторску дисертацију (максимално 1 страна)

Bennett, R.E. (1991). ON THE MEANINGS OF CONSTRUCTED RESPONSE. *ETS Research Report Series*, 1991.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.

Burrows, S., Gurevych, I. & Stein, B. The Eras and Trends of Automatic Short Answer Grading. *Int J Artif Intell Educ* 25, 60–117 (2015). <https://doi.org/10.1007/s40593-014-0026-8>

Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., ... & Wen, Q. (2025). Llm agents for education: Advances and applications. *arXiv preprint arXiv:2503.11733*, 2.

European Union. (2024). *Artificial Intelligence Act (Regulation (EU) 2024/1689)*.

- Fagbohun, O., Iduwe, N. P., Abdullahi, M., Ifaturoti, A., & Nwanna, O. M. (2024). Beyond traditional assessment: Exploring the impact of large language models on grading practices. *Journal of Artificial Intelligence, Machine Learning and Data Science*, 2(1), 1-8.
- Gao, M., Li, Y., Liu, B., Yu, Y., Wang, P., Lin, C. Y., & Lai, F. (2025). Single-agent or Multi-agent Systems? Why Not Both?. *arXiv preprint arXiv:2505.18286*.
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., ... & Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy & practice*, 4(3), 365-379.
- Idowu, J. A., & Almasoud, A. (2026). Specialists or Generalists? Multi-Agent and Single-Agent LLMs for Essay Grading. *arXiv preprint arXiv:2601.22386*.
- Jiang, Y. H., Li, R., Zhou, Y., Qi, C., Hu, H., Wei, Y., ... & Wu, Y. (2024). Ai agent for education: von neumann multi-agent system framework. *arXiv preprint arXiv:2501.00083*.
- Jiang, Y. H., Liu, T. Y., Zhuang, X., Hu, H., Li, R., & Jia, R. (2024). Enhancing educational practices with multi-agent systems: A review. *Enhancing educational practices: Strategies for assessing and improving learning outcomes*, 47-65.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2), 130-144.
- Kamalov, F., Calonge, D.S., Smail, L., Azizov, D., Thadani, D.R., Kwong, T., & Atif, A. (2025). Evolution of AI in Education: Agentic Workflows. *ArXiv, abs/2504.20082*.
- Lagakis, P., & Demetriadis, S. (2024, June). Evaai: a multi-agent framework leveraging large language models for enhanced automated grading. In *International conference on intelligent tutoring systems* (pp. 378-385). Cham: Springer Nature Switzerland.
- Li, G., Chen, L., Tang, C., Švábenský, V., Deguchi, D., Yamashita, T., & Shimada, A. (2025, June). Single-agent vs. multi-agent LLM strategies for automated student reflection assessment. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 300-311). Singapore: Springer Nature Singapore.
- Nitko, A. J., & Brookhart, S. M. (2006). *Educational assessment of students*. Prentice-Hall, Inc..
- Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation*. NY: Routledge.
- Wang, Y., Ding, Z., Wu, X., Sun, S., Liu, N., & Zhai, X. (2025). AutoSCORE: Enhancing Automated Scoring with Multi-Agent Large Language Models via Structured Component Recognition. *arXiv preprint arXiv:2509.21910*.
- Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., ... & Wang, C. (2024, August). Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First conference on language modeling*.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International journal of educational technology in higher education*, 16(1), 39.
- Zhang, R. (2025, July). Multi-Agent Systems for Learning Assessment in Education: A Comprehensive Survey. In *Proceedings of the 2025 3rd International Conference on Educational Knowledge and Informatization* (pp. 388-391).
- Zhang, X., Zhang, C., Sun, J., Xiao, J., Yang, Y., & Luo, Y. (2025). Eduplanner: Llm-based multi-agent systems for customized and intelligent instructional design. *IEEE Transactions on Learning Technologies*.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 1-124.

## 6. Методе које се користе у истраживању (максимално 1 страна)

Методологија истраживања биће усмерена на остваривање општег циља дисертације који укључује дизајн, развој и експерименталну евалуацију МАС-а заснованог на великим језичким моделима, за тачно, конзистентно и објашњиво аутоматско оцењивање студентских радова. У складу са тим, експерименти ће бити дизајнирани тако да омогуће проверу постављених хипотеза, као и остваривање наведених конкретних циљева дисертације.

У оквиру циља испитивања утицаја мултиагентског приступа на тачност оцењивања, а који се односи на хипотезу Х1, биће имплементирани базни, моноагентски систем, и МАС за аутоматско оцењивање, оба заснована на великим језичким моделима. Поређење ова два система вршиће се на јавно доступним скуповима података за оцењивање есеја и кратких одговора, као и на припремљеном скупу података за оцењивање програмског кода. Поред међусобног

поређења, резултати предложених система биће упоређени и са постојећим системима за аутоматско оцењивање који су на истим јавним скуповима података показали високе резултате, како би се у анализу укључили и приступи који нису засновани на агентским архитектурама.

Ради испитивања утицаја адаптивне специјализације агената на тачност аутоматског оцењивања, у складу са хипотезом Х2, биће имплементиран MAC са специјализованим агентима. Агенти ће бити специјализовани према појединачним критеријумима оцењивања дефинисаним у рубрици или утврђених претходним прегледом релевантне литературе. Адаптивност MAC-а са специјализацијом биће омогућена увођењем оркестратор агента који ће, на основу анализе рубрике за оцењивање за дати тип задатка, утврдити по којим критеријумима је потребно оценити студентски рад и на основу њих иницијализовати одговарајући број специјализованих агената. Поређење ће се вршити са MAC-ом без адаптивне специјализације агената, над истим скуповима података, чиме ће се утврдити у којој мери специјализација агената у односу на тип задатка утиче на тачност аутоматског оцењивања.

Како би се испитао утицај начина интеграције релевантног спољног знања на тачност оцењивања, у оквиру хипотезе Х3, биће имплементирани различити приступи коришћења спољног знања у оквиру MAC-а. У ту сврху биће коришћени разни извори знања, која могу бити општи за дати задатак (нпр. курикулум, уџбеник, референтни одговор) или специфични за критеријум оцењивања (нпр. дефиниције појмова, доменско знање). Код општег начина коришћења спољног знања, сви агенти ће користити заједнични општи контекст прибављен из општих извора знања. Код специфичног начина, агенти ће прибављати и користити знање релевантно свом критеријуму оцењивања. Поред тога, биће испитан и хибридни приступ, у оквиру ког ће агенти користити комбинацију општег и специфичног знања. Поређењем перформанси ових система међусобно, као и са верзијом MAC-а без приступа спољном знању, биће анализирано у којој мери интеграција и различити начини интеграције релевантног спољног знања утичу на тачност аутоматског оцењивања.

У циљу испитивања утицаја начина агрегације оцена појединачних агената у MAC-у на тачност аутоматског оцењивања, у оквиру хипотезе Х4, биће упоређене две варијанте MAC. У првој варијанти, коначна оцена ће се формирати применом једноставних облика агрегације (просека, тежинског просека) оцена које генеришу појединачни агенти. У другој варијанти, у те сврхе биће уведен агент за агрегацију, који ће анализирати оцене појединачних агената и њихових образложења, и на основу њих просудити коначну оцену студентског рада. Поређењем ових варијанти MAC-а над истим скуповима података утврдиће се колики утицај агентски вид агрегације оцена има на тачност аутоматског оцењивања.

За евалуацију тачности свих претходно наведених система, између осталог, користиће се корелација (Pearson) као мера линеарне повезаности са референтним оценама, Mean Absolute Error и Root Mean Squared Error као мере просечног одступања системских од референтних оцена, као и Quadratic Weighted Kappa (QWK) која мери степен сагласности између системских и референтних оцена.

У циљу испитивања утицаја MAC-а на транспарентност и објашњивост процеса аутоматског оцењивања, а у оквиру хипотезе Х5, биће упоређени мултиагентски и моноагентски системи засновани на великим језичким моделима, који генеришу образложења својих одлука у вези са процесом оцењивања. Квалитет генерисаних образложења оцена биће процењен различитим методама у зависности од доступности референтног образложења у скупу података. Код скупова података где није доступно референтно образложење, биће примењен приступ у ком ће независан језички модел евалуирати структуру, утемељеност и аргументованост генерисаних образложења у односу на задатак и критеријуме оцењивања, применом *LLM-as-a-judge* приступа. Код скупова података где ће бити доступна референтна наставничка образложења оцена, квалитет генерисаних образложења ће бити процењен применом метрика сличности са референтним образложењем, као што су ROUGE и BERTScore.

Сви експерименти биће реализовани у програмском језику Пајтон (Python). За имплементацију MAC-а користиће се библиотеке за велике језичке моделе и агентски оквири (LangChain, LangGraph и сл). За имплементацију приступа спољном знању користиће се библиотеке за рад са векторским базама (FAISS, ChromaDB, Qdrant или сл), као и векторски (embedding) модели доступни путем платформи као што је *HuggingFace*. За обраду података користиће се библиотеке попут *pandas* и *NumPy*. Велики језички модели биће хостовани коришћењем платформе *Ollama* и сл. С обзиром на брз развој нових технологија у области вештачке интелигенције и MAC-а, предложене технологије ће бити подложне промени, те је могуће да ће у току израде дисертације бити коришћени и други алати, библиотеке и платформе које у датом тренутку буду актуелне и погодније за реализацију експеримената, уз задржавање постављене методологије и експерименталног дизајна.

## 7. Мултидисциплинарност теме (максимално 1/2 стране)

Мултидисциплинарни карактер ове теме огледа се у повезивању више научних области. Из угла рачунарских наука, истраживање се бави применом савремених метода вештачке интелигенције, пре свега применом великих језичких

модела из области процесирања природних језика, метода развоја MAC-а, као и приступа из области софтверског инжењерства. Истовремено, рад се ослања на педагошке приступе вредновања знања, посебно кроз употребу рубрика за оцењивање у аутоматским системима. Додатно, постоји веза и са едукативном психологијом у домену транспарентности и објашњивости процеса оцењивања, који утичу на поверење наставника и студената у аутоматизоване системе, као и на перцепцију правичности и поузданости процеса аутоматског оцењивања. Повезаност са педагогијом и едукативном психологијом огледа се и кроз примену концепта повратне информације (feedback), који омогућује да аутоматски систем пружи аргументовано образложење додељене оцене студенту и наставнику. Истраживање се може посматрати и у ширем контексту области образовних технологија, јер доприноси развоју интелигентних система намењених подршци процесу наставе и вредновању знања.

## **8. Очекивани научни допринос докторске дисертације**

Очекивани научни допринос ове докторске дисертације обухвата дизајн, развој и мултиаспектну евалуацију иновативног приступа објашњивом аутоматском оцењивању студентских радова, заснованом на систему интелигентних агената, који интегрише велике језичке моделе и спољне изворе знања ради унапређења тачности, конзистентности и објашњивости процеса аутоматског оцењивања.

Конкретно, очекивани научни доприноси се огледају у следећем:

- Развој архитектуре MAC-а заснованог на великим језичким моделима за аутоматско оцењивање студентских радова, са приступом спољном знању, који омогућује адаптивну специјализацију агената у односу на тип задатка и по различитим критеријумима оцењивања дефинисаних рубриком за оцењивање, као и механизам за агрегацију појединачних процена агената у јединствену коначну оцену, уз образложење додељене оцене
- Емпиријски утврђен утицај мултиагентског приступа на тачност аутоматског оцењивања у односу на постојеће моноагентске и неагентске приступе, чиме се доприноси разумевању потенцијала примене MAC-а у образовним окружењима
- Емпиријски процењен утицај адаптивне специјализације агената у односу на тип задатка и по појединачним критеријумима оцењивања на тачност аутоматског оцењивања
- Емпиријски процењен утицај различитих начина интеграције спољног знања у MAC на тачност аутоматског оцењивања
- Емпиријски утврђен утицај различитих начина агрегације оцена појединачних агената на коначну оцену MAC-а
- Приступ за транспарентно и објашњиво аутоматско оцењивање кроз генерисање образложења оцене студентског рада, засновано на рубрикама и релевантном спољном знању
- Предлог експерименталног оквира за систематско испитивање MAC-а заснованих на великим језичким моделима, у задацима аутоматског оцењивања студентских радова, који омогућује контролисано поређење различитих варијанти система и конфигурација агената
- Допринос бољем разумевању примене великих језичких модела и мултиагентских архитектура у образовању, као и њихових предности и ограничења у задацима аутоматског оцењивања и пружања повратних информација
- Могућност примене предложеног MAC-а у реалном образовном окружењу, посебно у задатку оцењивања програмског кода

## **9. Изјава да предложену тему кандидат није пријављивао на другој високошколској установи у земљи или иностранству**

Ја, Марко Путниковић, изјављујем да предложену тему докторске дисертације „Мултиагентски приступи засновани на великим језичким моделима за објашњиво аутоматско оцењивање студентских радова” нисам пријављивао на другој високошколској установи у земљи или иностранству.

## **10. Предлог два ментора и комисије за оцену теме докторске дисертације (име, презиме, звање, институција, ужа научна област)**

Предлог ментора:

1. Јелена Јовановић, редовни професор, Факултет организационих наука, Универзитет у Београду, ужа научна област: софтверско инжењерство
2. Александар Бауцал, редовни професор, Филозофски факултет, Универзитет у Београду, ужа научна област: општа психологија

Предлог комисије:

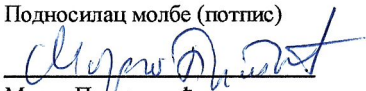
1. Ранка Станковић, редовни професор, Рударско-геолошки факултет, Универзитет у Београду, ужа научна област: математика и информатика
2. Срећко Јоксимовић, професор, *School of Education, Adelaide University*, ужа научна област: аналитике учења, вештачка интелигенција у образовању
3. Владан Девеџић, редовни професор, Факултет организационих наука, Универзитет у Београду, ужа научна област: софтверско инжењерство

**11. Прилози:**

1. Сагласност ментора (треба да садржи име, презиме, звање, институцију и потпис)
2. Подаци о ментору
3. Мишљење одговарајућих етичких комитета о етичким аспектима предложене теме (клиничка испитивања или експерименти на животињама)

У Београду, датум 13-3-2022.

Подносилац молбе (потпис)

  
Марко Путниковић

**Form 1 – Doctoral Dissertation Topic Proposal**  
**(The candidate submits the proposal to the Program Council)**  
Study program: Intelligent Systems

Candidate: Marko Putniković

I kindly request approval to prepare a doctoral dissertation under the title:

*"Multi-Agent Approaches Based on Large Language Models for Explainable Automatic Grading of Student Work"*

### **1. Candidate Biography**

Marko Putniković was born on May 13, 1993 in Valjevo, Serbia. He enrolled in undergraduate studies at the Faculty of Organizational Sciences, University of Belgrade in 2013, in the program Information Systems and Technologies. He graduated in 2017, earning the degree of Engineer of Organizational Sciences – Information Systems and Technologies. He enrolled in master's studies in 2017 at the University of Grenoble, in the program Artificial Intelligence and Semantic Web, as a scholarship recipient of the French government. He completed his master's studies in 2018, earning the degree of Master Informatics. During his master's studies, he completed a five-month internship (stage) at the Computer Science Laboratory of the University of Grenoble. He enrolled in doctoral studies in 2019 at the University of Belgrade, in the multidisciplinary study program at the university, specialization Intelligent Systems. During his studies, he worked on the application of natural language processing methods in education and published a research paper in an M21 journal on the application of these methods in grading student work (2.1.1, 2.2.1).

Concurrently with his doctoral studies, the candidate worked at NCR as a software engineer from 2019 to 2022, after which he moved to Robert Bosch where he continues to work as an artificial intelligence engineer. At Robert Bosch, he worked on an innovation research project on the application of natural language processing methods for web data extraction, within which he published a research paper at a conference (2.2.2).

### **2. Candidate Bibliography (categorized according to the classification of the relevant Ministry)**

#### **2.1. Published papers or accepted for publication (DOI number or editor's statement)**

*Distinguished international journals (M21):*

1. Putnikovic, M., & Jovanovic, J. (2023). Embeddings for automatic short answer grading: A scoping review. *IEEE Transactions on Learning Technologies*, 16(2), 219-231.

#### **2.2. Conferences and gatherings**

1. Putnikovic, M., & Jovanovic, J. (2022). Word Embeddings for Automatic Short Answer Grading (pp. 197-213). World Scientific Publishing Co. Pte. Ltd.
2. Hätyy, A., Milchevski, D., Döring, K., Putnikovic, M., Mesgar, M., Novović, F., ... & Stranjanac, I. (2024, November). A Cost-Efficient Modular Sieve for Extracting Product Information from Company Websites. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track* (pp. 1444-1456).
3. Stanković, R., Rađenović, J., Škorić, M., & Putnikovic, M. (2025, March 27). Learning Word Embeddings using Lexical Resources and Corpora. *ICIST – International Conference on Information Society and Technology (ICIST 2025)*, Kopaonik, Serbia. Zenodo. <https://doi.org/10.5281/zenodo.15093900>

### **3. Subject and Objectives of the Doctoral Dissertation**

#### **3.1 Subject of the Doctoral Dissertation (maximum 1 page)**

The subject of this doctoral dissertation is the automatic grading of student work and the provision of explanations about the grading process through the application of multi-agent systems based on large language models.

Grading student work represents one of the main processes within the educational system, as it enables assessment and provides feedback on the extent of student knowledge acquisition. Assessment can be summative in nature, which involves assigning a grade and provides a formal basis for advancement and obtaining an academic degree; and formative in nature, which includes providing feedback to the student, monitoring learning progress, and guiding the further development of student competencies. For the instructor, the grading task is usually time-constrained and cognitively demanding depending on the number of students, which can lead to fatigue, affect consistency, and reduce the possibility of better student support.

Additionally, instructor subjectivity, namely personal beliefs and expectations, reliance on prior experience and intuition, can lead to inconsistency in grading and affect outcomes. With the development of online forms of education, automatic grading systems are gaining increasing importance, with the purpose of enabling faster and more consistent grading of large numbers of student papers, as well as relieving instructors of the part of the grading process and providing justification for assigned grades that can be automated.

Formally, the tasks that are graded can be divided into closed (e.g., multiple-choice tasks) and open-ended (constructed response) (Bennett, 1991). Closed-ended tasks involve the selection of one or more offered answers, where grading is deterministic, i.e., based on a predefined solution. On the other hand, open-ended tasks can be divided into those with short and long free response text (essays), but depending on the subject domain, they can also be programming code, mathematical proofs, etc. For the open-ended type, which will be the focus of this doctoral dissertation research, content processing and analysis are required, which requires the application of more sophisticated natural language processing methods.

Historically, methods for grading open-ended tasks can be divided into those based on (1) templates and rules, (2) statistical text characteristics, and (3) machine and deep learning (Burrows et al. 2015). The deep learning approach has been particularly prevalent in recent years due to the significant success of applying large language models to grading student work (Fagbohun et al. 2024). However, in a large number of existing studies, all tasks within the grading process are performed by a single large language model, which can lead to various problems generally attributed to such models, such as hallucinations and loss of context. Moreover, grading with a single large language model can lead to the mixing of grading criteria when it is necessary to evaluate different aspects of the work separately. Additionally, the information that a single large language model possesses may be outdated or irrelevant, and it is necessary to enrich the prompt with current knowledge relevant to the grading context.

An approach that could improve the grading process and the provision of explanations of this process using large language models is the application of grading systems based on multiple mutually coupled large language models. Such systems are called multi-agent systems (MAS), where an agent is a software component that, among other things, uses a large language model to perform a specific task in the grading process (Kamalov et al. 2025). In addition to large language models, an agent can use various other software tools to enrich its current knowledge about the relevant context and produce better results (web search, data analysis libraries, database access, etc.). The application of MAS based on large language models in automatic grading still represents a relatively new and insufficiently systematically researched area. In particular, there is a lack of studies analyzing the impact of architecture, i.e., the structural organization of MAS, on grading results. Specifically, there is a limited number of studies examining the advantages and disadvantages of multi-agent approaches compared to other automatic grading systems, examining the impact of agent specialization across different aspects of grading on grading accuracy, methods of aggregating individual agent grades into a final grade, and analyzing the effects of integrating external knowledge into MAS. Furthermore, existing research insufficiently addresses the question of transparency and explainability of the grading process in MAS. Justification of the grade assigned by an automatic system is necessary to ensure grading transparency and enable verification of the application of grading rubrics, thereby building trust in the system. Additionally, regulatory legal frameworks such as the European Union Artificial Intelligence Act (European Union, 2024, Article 86) require a certain level of transparency in automatic decision-making, especially in high-risk application domains, which include education.

### **3.2. Objectives of the Doctoral Dissertation (maximum 1 page)**

The general objective of this doctoral dissertation is the development of an accurate, consistent, and explainable MAS based on large language models for the automatic grading of student work. In accordance with the defined general objective, the following specific objectives have been formulated:

1. Examine the extent to which agent specialization relative to the rubric contributes to the accuracy of automatic grading
2. Examine the contribution of integrating relevant external knowledge (e.g., curriculum or instructional materials) to the accuracy of automatic grading
3. Analyze how different methods of aggregating individual agent assessments affect the accuracy of automatic grading
4. Ensure transparency and explainability of the MAS grading process through the generation of grounded and well-argued explanations

### **3.3 Hypotheses**

Based on the analysis of relevant and available literature, as well as identified under-researched topics in the field of automatic grading, we formulate the following hypotheses:

- H1. The application of MAS in automatic grading of student work leads to the increase in accuracy of automatic grading compared to existing approaches.
- H2. Adaptive specialization of agents relative to the task type and based on the grading rubric, leads to the increase in accuracy of automatic grading compared to a multi-agent approach without adaptive agent specialization.
- H3. Integration of relevant external knowledge into MAS, determined by the type of task being graded, leads to greater accuracy of automatic grading compared to MAS without the use of external knowledge.
- H4. Introduction of an aggregation agent into MAS, which makes a final grade based on individual agent assessments, leads to greater accuracy of automatic grading compared to MAS in which the final grade is obtained through simple aggregation (e.g., averaging) of individual agent grades.
- H5. A multi-agent system based on large language models enables greater transparency and explainability of the automatic grading process compared to a single-agent approach, through the generation of justifications for the assigned grade.

#### 4. Work Plan

The work plan for the doctoral dissertation will consist of the following steps:

- Review of relevant literature in the fields of automatic grading, large language models, and MAS
- Analysis of existing applications of MAS based on large language models for automatic grading
- Selection of publicly available datasets for the evaluation of automatic grading of essays and short answers, as well as preparation of a dataset for grading programming code
- Collection of grading rubrics by task type through a review of relevant literature and analysis of rubrics from publicly available datasets
- Design of a single-agent system for automatic grading (baseline model)
- Design of MAS with adaptive agent specialization
- Design of MAS with a grade aggregation agent
- Design of MAS with access to external knowledge
- Incorporation of mechanisms for generating grounded and well-reasoned grade justifications within MAS to ensure transparency and explainability of the grading process
- Selection of representative large language models to be used in experimental evaluation
- Definition of evaluation metrics, i.e., assessment of the accuracy of the designed MAS, as well as assessment of the quality of automatically generated grade justifications
- Implementation of proposed systems and experimental setups
- Conducting experiments and collecting results
- Analysis of results, visualization and interpretation in relation to the stated hypotheses
- Discussion of research limitations and proposal of future directions for development

#### 5. Bibliographic Data Relevant to the Doctoral Dissertation (maximum 1 page)

Bennett, R.E. (1991). ON THE MEANINGS OF CONSTRUCTED RESPONSE. *ETS Research Report Series*, 1991.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.

Burrows, S., Gurevych, I. & Stein, B. The Eras and Trends of Automatic Short Answer Grading. *Int J Artif Intell Educ* 25, 60–117 (2015). <https://doi.org/10.1007/s40593-014-0026-8>

Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., ... & Wen, Q. (2025). Llm agents for education: Advances and applications. *arXiv preprint arXiv:2503.11733*, 2.

European Union. (2024). *Artificial Intelligence Act (Regulation (EU) 2024/1689)*.

Fagbohun, O., Iduwe, N. P., Abdullahi, M., Ifaturoti, A., & Nwanna, O. M. (2024). Beyond traditional assessment: Exploring the impact of large language models on grading practices. *Journal of Artificial Intelligence, Machine Learning and Data Science*, 2(1), 1-8.

Gao, M., Li, Y., Liu, B., Yu, Y., Wang, P., Lin, C. Y., & Lai, F. (2025). Single-agent or Multi-agent Systems? Why Not Both?. *arXiv preprint arXiv:2505.18286*.

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., ... & Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*.

- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy & practice*, 4(3), 365-379.
- Iadow, J. A., & Almasoud, A. (2026). Specialists or Generalists? Multi-Agent and Single-Agent LLMs for Essay Grading. *arXiv preprint arXiv:2601.22386*.
- Jiang, Y. H., Li, R., Zhou, Y., Qi, C., Hu, H., Wei, Y., ... & Wu, Y. (2024). Ai agent for education: von neumann multi-agent system framework. *arXiv preprint arXiv:2501.00083*.
- Jiang, Y. H., Liu, T. Y., Zhuang, X., Hu, H., Li, R., & Jia, R. (2024). Enhancing educational practices with multi-agent systems: A review. *Enhancing educational practices: Strategies for assessing and improving learning outcomes*, 47-65.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2), 130-144.
- Kamalov, F., Calonge, D.S., Smail, L., Azizov, D., Thadani, D.R., Kwong, T., & Atif, A. (2025). Evolution of AI in Education: Agentic Workflows. *ArXiv, abs/2504.20082*.
- Lagakis, P., & Demetriadis, S. (2024, June). Evaai: a multi-agent framework leveraging large language models for enhanced automated grading. In *International conference on intelligent tutoring systems* (pp. 378-385). Cham: Springer Nature Switzerland.
- Li, G., Chen, L., Tang, C., Švábenský, V., Deguchi, D., Yamashita, T., & Shimada, A. (2025, June). Single-agent vs. multi-agent LLM strategies for automated student reflection assessment. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 300-311). Singapore: Springer Nature Singapore.
- Nitko, A. J., & Brookhart, S. M. (2006). *Educational assessment of students*. Prentice-Hall, Inc..
- Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation*. NY: Routledge.
- Wang, Y., Ding, Z., Wu, X., Sun, S., Liu, N., & Zhai, X. (2025). AutoSCORE: Enhancing Automated Scoring with Multi-Agent Large Language Models via Structured Component Recognition. *arXiv preprint arXiv:2509.21910*.
- Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., ... & Wang, C. (2024, August). Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First conference on language modeling*.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International journal of educational technology in higher education*, 16(1), 39.
- Zhang, R. (2025, July). Multi-Agent Systems for Learning Assessment in Education: A Comprehensive Survey. In *Proceedings of the 2025 3rd International Conference on Educational Knowledge and Informatization* (pp. 388-391).
- Zhang, X., Zhang, C., Sun, J., Xiao, J., Yang, Y., & Luo, Y. (2025). Eduplanner: Llm-based multi-agent systems for customized and intelligent instructional design. *IEEE Transactions on Learning Technologies*.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 1-124.

## 6. Research Methods (maximum 1 page)

The research methodology will be directed toward achieving the general objective of the dissertation, which includes the design, development, and experimental evaluation of a MAS based on large language models for accurate, consistent, and explainable automatic grading of student work. Accordingly, experiments will be designed to enable the verification of the stated hypotheses, as well as the achievement of the specified concrete objectives of the dissertation.

Within the objective of examining the impact of the multi-agent approach on grading accuracy, which relates to hypothesis H1, a baseline single-agent system and a MAS for automatic grading will be implemented, both based on large language models. The comparison of these two systems will be conducted on publicly available datasets for grading essays and short answers, as well as on a prepared dataset for grading programming code. In addition to mutual comparison, the results of the proposed systems will also be compared with existing automatic grading systems that have demonstrated high performance on the same public datasets, in order to include approaches that are not based on agent architectures in the analysis.

To examine the impact of adaptive agent specialization on the accuracy of automatic grading, in accordance with hypothesis H2, a MAS with specialized agents will be implemented. Agents will be specialized according to individual grading criteria defined in the rubric or established through a prior review of relevant literature. The adaptability of the MAS with specialization will be enabled by introducing an orchestrator agent that will, based on the analysis of the grading rubric for a given task type, determine which criteria need to be applied to grade the student work and initialize the corresponding

number of specialized agents accordingly. The comparison will be made with a MAS without adaptive agent specialization, over the same datasets, thereby determining the extent to which agent specialization relative to the task type affects the accuracy of automatic grading.

To examine the impact of different methods of integrating relevant external knowledge on grading accuracy, within hypothesis H3, different approaches to using external knowledge within MAS will be implemented. For this purpose, various sources of knowledge will be used, which can be general for a given task (e.g., curriculum, textbook, reference answer) or specific to the grading criteria (e.g., definitions of terms, domain knowledge). In the general approach to using external knowledge, all agents will use a shared general context obtained from general knowledge sources. In the specific approach, agents will obtain and use knowledge relevant to their grading criterion. Additionally, a hybrid approach will be examined, in which agents will use a combination of general and specific knowledge. By comparing the performance of these systems with each other, as well as with a version of MAS without access to external knowledge, we will analyze the extent to which the integration and different methods of integration of relevant external knowledge affect the accuracy of automatic grading.

To examine the impact of different methods of aggregating individual agent grades in MAS on the accuracy of automatic grading, within hypothesis H4, two variants of MAS will be compared. In the first variant, the final grade will be formed by applying simple forms of aggregation (average, weighted average) of grades generated by individual agents. In the second variant, an aggregation agent will be introduced for this purpose, which will analyze the grades of individual agents and their grade justifications, and based on them determine the final grade of the student work. By comparing these MAS variants over the same datasets, we will determine the impact of the agent-based grade aggregation on the accuracy of automatic grading.

For the evaluation of the accuracy of all the aforementioned systems, among other measures, Pearson correlation will be used as a measure of linear association with reference grades, Mean Absolute Error and Root Mean Squared Error as measures of average deviation of system grades from reference grades, as well as Quadratic Weighted Kappa (QWK), which measures the degree of agreement between system and reference grades.

To examine the impact of MAS on the transparency and explainability of the automatic grading process, within hypothesis H5, we will compare multi-agent and single-agent systems based on large language models that generate justifications for their grading decisions. The quality of the generated grade justifications will be assessed by different methods depending on the availability of reference justifications in the dataset. For datasets where reference justifications are not available, we will apply an approach in which an independent language model will evaluate the structure, groundedness, and argumentation of the generated justifications in relation to the task and grading criteria. For that purpose, we will use the LLM-as-a-judge approach. For datasets where reference teacher justifications will be available, the quality of the generated justifications will be assessed using metrics that measure similarity with the reference justification, such as ROUGE and BERTScore.

All experiments will be implemented in the Python programming language. For the implementation of MAS, we will use libraries for large language models and agent frameworks (LangChain, LangGraph, etc.). For the implementation of external knowledge access, we will use libraries for working with vector databases (FAISS, ChromaDB, Qdrant, or similar), as well as vector (embedding) models available through platforms such as HuggingFace. For data processing, libraries such as pandas and NumPy will be used. Large language models will be hosted using the Ollama platform and similar. Given the rapid development of new technologies in the field of artificial intelligence and MAS, the proposed technologies will be subject to change, and it is possible that during the preparation of the dissertation, other tools, libraries, and platforms that are current and more suitable for implementing the experiments at the given time will be used, while maintaining the established methodology and experimental design.

## **7. Multidisciplinarity of the Topic (maximum 1/2 page)**

The multidisciplinary character of this topic is reflected in the connection of multiple scientific fields. From the perspective of computer science, the research deals with the application of modern artificial intelligence methods, primarily the application of large language models from the field of natural language processing, methods of MAS development, as well as approaches from the field of software engineering. At the same time, the work draws on pedagogical approaches to knowledge evaluation, particularly through the use of grading rubrics in automated systems. Additionally, there is a connection with educational psychology in the domain of transparency and explainability of the grading process, which affects the trust of instructors and students in automated systems, as well as the perception of fairness and reliability of the automatic grading process. The connection with pedagogy and educational psychology is also reflected through the application of the feedback, which enables the automatic system to provide a well-argued justification of the assigned grade to both the student and the instructor. The research can also be viewed in the broader context of the field of educational technologies, as it contributes to the development of intelligent systems designed to support the teaching process and knowledge evaluation.

## 8. Expected Scientific Contribution of the Doctoral Dissertation

The expected scientific contribution of this doctoral dissertation encompasses the design, development, and multi-aspect evaluation of an innovative approach to explainable automatic grading of student work, based on a system of intelligent agents that integrates large language models and external knowledge sources to improve the accuracy, consistency, and explainability of the automatic grading process.

Specifically, the expected scientific contributions are reflected in the following:

- Development of a MAS architecture based on large language models for automatic grading of student work, with access to external knowledge, enabling adaptive specialization of agents relative to the task type and across different grading criteria defined by the grading rubric, as well as a mechanism for aggregating individual agent assessments into a single final grade, along with justification of the assigned grade
- Empirically established impact of the multi-agent approach on the accuracy of automatic grading compared to existing single-agent and non-agent approaches, thereby contributing to understanding the potential of MAS application in educational settings
- Empirically assessed impact of adaptive agent specialization relative to the task type and across individual grading criteria on the accuracy of automatic grading
- Empirically assessed impact of different methods of integrating external knowledge into MAS on the accuracy of automatic grading
- Empirically established impact of different methods of aggregating individual agent grades on the final MAS grade
- An approach for transparent and explainable automatic grading through the generation of grade justifications for student work, based on rubrics and relevant external knowledge
- Proposal of an experimental framework for the systematic examination of MAS based on large language models in automatic grading tasks of student work, enabling controlled comparison of different system variants and agent configurations
- Contribution to a better understanding of the application of large language models and multi-agent architectures in education, as well as their advantages and limitations in automatic grading tasks and provision of feedback
- Potential for application of the proposed MAS in a real educational setting, particularly in the task of grading programming code

## 9. Declaration that the proposed topic has not been submitted at another higher education institution in the country or abroad

I, Marko Putniković, hereby declare that I have not submitted the proposed doctoral dissertation topic "Multi-Agent Approaches Based on Large Language Models for Explainable Automatic Grading of Student Work" at any other higher education institution in the country or abroad.

## 10. Proposal for two mentors and a committee for evaluating the doctoral dissertation topic

*(name, surname, title, institution, narrow scientific field)*

### Proposed mentors:

1. Jelena Jovanović, Full Professor, Faculty of Organizational Sciences, University of Belgrade, scientific field: software engineering
2. Aleksandar Baucal, Full Professor, Faculty of Philosophy, University of Belgrade, scientific field: general psychology

### Proposed committee:

1. Ranka Stanković, Full Professor, Faculty of Mining and Geology, University of Belgrade, scientific field: mathematics and computer science
2. Srećko Joksimović, Professor, School of Education, Adelaide University, scientific field: learning analytics, artificial intelligence in education
3. Vladan Devedžić, Full Professor, Faculty of Organizational Sciences, University of Belgrade, scientific field: software engineering

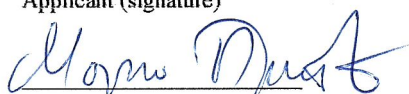
**11. Appendices:**

1. Mentor's consent (should contain name, surname, title, institution, and signature)
2. Mentor's data
3. Opinion of the relevant ethics committees on the ethical aspects of the proposed topic (clinical trials or experiments on animals)

In Belgrade, date

13.3.2026.

Applicant (signature)



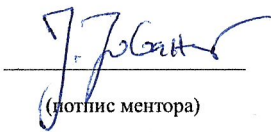
Marko Putniković

Прилог уз Образац 1

Универзитет у Београду

САГЛАСНОСТ МЕНТОРА

Ја, Јелена Јовановић, редовни професор на Факултету организационих наука Универзитета у Београду, сагласна сам са темом докторске дисертације **Мултиагентски приступи засновани на великим језичким моделима за објашњиво аутоматско оцењивање студентских радова**, кандидата Марка Путниковића, бр. индекса 2/2019, студијски програм Интелигентни системи.



(потпис ментора)

У Београду,

2.3.2026.

ПОДАЦИ О МЕНТОРУ

Име и презиме: Јелена Јовановић

Звање: Редовни професор

Списак радова објављених у научним часописима са Science Citation Index (SCI) листе који квалификују ментора за вођење докторске дисертације:

1. Khosravi, H., Shibani, A., Jovanovic, J., Pardos, Z. A., & Yan, L. (2025). Generative AI and Learning Analytics: Pushing Boundaries, Preserving Principles. *Journal of Learning Analytics*, 12(1), Article 1. <https://doi.org/10.18608/jla.2025.8961>
2. Jovanović, J., Zamecnik, A., Barthakur, A., & Dawson, S. (2024). Curriculum analytics: Exploring assessment objectives, types, and grades in a study program. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-13015-0>
3. Liu, Q., Shakya, R., Jovanovic, J., Khalil, M., & de la Hoz-Ruiz, J. (2025). Ensuring privacy through synthetic data generation in education. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13576>
4. Jovanovic, J., Gašević, D., Yan, L., Baker, G., Murray, A., & Gasevic, D. (2024). Explaining trace-based learner profiles with self-reports: The added value of psychological networks. *Journal of Computer Assisted Learning*, 40(4), 1481–1499. <https://doi.org/10.1111/jcal.12968>
5. Putnikovic, M., & Jovanovic, J. (2023). Embeddings for Automatic Short Answer Grading: A Scoping Review. *IEEE Transactions on Learning Technologies*, 1–13. <https://doi.org/10.1109/TLT.2023.3253071>

(Павести најмање један рад у случају менторства дисертације која се пријављује до 31. 12. 2008. године, односно најмање три рада у случају менторства дисертације која се пријављује у периоду од 1. 1. 2009. до 31. 12. 2009. године, односно најмање пет радова у случају менторства дисертације која се пријављује почев од 1. 1. 2010. године.)

## Прилог уз Образац 1

Универзитет у Београду

### САГЛАСНОСТ МЕНТОРА

Ја, Александар Бауцал, редовни професор на Филозофском факултету Универзитета у Београду, сагласан сам са темом докторске дисертације **Мултиагентски приступи засновани на великим језичким моделима за објашњиво аутоматско оцењивање студентских радова**, кандидата Марка Путниковића, бр. индекса 2/2019, студијски програм Интелигентни системи.



(потпис ментора)

У Београду,

12.3.2026.

### ПОДАЦИ О МЕНТОРУ

Име и презиме: Александар Бауцал

Звање: Редовни професор

Списак радова објављених у научним часописима са Science Citation Index (SCI) листе који квалификују ментора за вођење докторске дисертације:

1. Pedaste, M., Kallas, K., & Baucal, A. (2023). Digital competence test for learning in schools: Development of items and scales. *Computers & Education*, 203, 104830. <https://doi.org/10.1016/j.compedu.2023.104830>
2. Radišić, J., Videnović, M., & Baucal, A. (2015). Math anxiety—contributing school and individual level factors. *European Journal of Psychology of Education*, 30(1), 1-20. <https://doi.org/10.1007/s10212-014-0224-7>
3. Leijen, Ä., Pedaste, M., & Baucal, A. (2022). Assessing student teachers' agency and using it for predicting commitment to teaching. *European Journal of Teacher Education*, 45(5), 600-616. <https://doi.org/10.1080/02619768.2021.1889507>
4. Leijen, Ä., Baucal, A., Pikk, K., Uibu, K., Pajula, L., & Sõrmus, M. (2024). Opportunities to develop student's math-related agency in primary education: The role of teacher beliefs. *European Journal of Psychology of Education*, 39(2), 1637-1659. <https://doi.org/10.1007/s10212-023-00771-9>
5. Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: development of items and scales. *International Journal of STEM Education*, 8(1), 19. <https://doi.org/10.1186/s40594-021-00278-z>

(Навести најмање један рад у случају менторства дисертације која се пријављује до 31. 12. 2008. године, односно најмање три рада у случају менторства дисертације која се пријављује у периоду од 1. 1. 2009. до 31. 12. 2009. године, односно најмање пет радова у случају менторства дисертације која се пријављује почев од 1. 1. 2010. године.)

Прилог уз Образац 1

Универзитет у Београду

### МОЛБА ЗА ИМЕНОВАЊЕ КОМИСИЈЕ

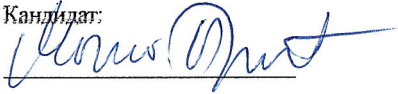
Молим Веће за студије при Универзитету у Београду да именује комисију за оцену научне заснованости теме докторске дисертације кандидата Марка Путниковића, под насловом **Мултиагентски приступи засновани на великим језичким моделима за објашњиво аутоматско оцењивање студентских радова**.

Предлаже се комисија за оцену научне заснованости теме докторске дисертације у следећем саставу:

1. Ранка Станковић, редовни професор, Рударско-геолошки факултет, Универзитет у Београду, ужа научна област: математика и информатика
2. Срећко Јоксимовић, професор, *School of Education, Adelaide University*, ужа научна област: аналитике учења, вештачка интелигенција у образовању
3. Владан Девеџић, редовни професор, Факултет организационих наука, Универзитет у Београду, ужа научна област: софтверско инжењерство

У Београду, 13.3.2026.

Кандидат:



Марко Путниковић