

КОМИСИЈИ ЗА СТУДИЈЕ II СТЕПЕНА ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА У БЕОГРАДУ

Комисија за студије II степена, Електротехничког факултета у Београду, на својој седници одржаној 28.07.2026 године именовала нас је у Комисију за преглед и оцену мастер рада дипл. инж. Елхан Даздаревић под насловом „Имплементација мултимодалног језичког модела за генерисање описа слике“. Након прегледа материјала Комисија подноси следећи

ИЗВЕШТАЈ

1. Биографски подаци кандидата

Елхан Даздаревић је рођен 15.03.2002. године у Новом Пазару. Основну школу и гимназију завршио је као носилац диплома „Вук Караџић“ и „Мухамед Хеваи Ускуфи“. У току основне школе био је проглашен за ученика генерације. Електротехнички факултет уписао је 2021. године. Дипломирао је на одсеку за Телекомуникације и информационе технологије, смер Информационо-комуникационе технологије, са просечном оценом 8,96. Дипломски рад одбранио је октобру 2025. године са оценом 10. Мастер студије на Електротехничком факултету у Београду, на Модулу за информационо-комуникационе технологије уписао је у новембру 2025. године.

2. Извештај о студијском истраживачком раду

Кандидат Елхан Даздаревић је као припрему за израду мастер рада урадио истраживање које се тиче релевантне литературе области којој припада тема мастер рада. Конкретно, анализирана је литература која се бави мултимедијалним системима, концептом обраде више различитих модалитета података, као и савременим архитектурама попут мултимодалних великих језичких модела (Multimodal Large Language Model - MLLM) и визуелно-језичких модела (Vision-Language Model - VLM). Истраживао је постојеће изазове у домену генеративног описивања визуелног садржаја помоћу вештачке интелигенције (Artificial Intelligence - AI), што је указало на неопходност даљег испитивања спајања визуелних и текстуалних репрезентација у оквиру кодер-декодер архитектуре.

3. Опис мастер рада

Мастер рад обухвата 34 стране, са укупно 13 слика, 3 табеле, 4 исечака програмског кода и 30 референци. Рад садржи увод, 4 поглавља и закључак (укупно 6 поглавља), као и списак литературе, списак скраћеница, списак слика, списак табела и списак исечака кода.

Прво поглавље представља увод у коме су описани предмет истраживања и циљ рада, уз објашњење основних појмова спајања визуелних и текстуалних података.

У другом поглављу је дат преглед литературе из области мултимодалних великих језичких модела, односно визуелно-језичких модела.

Треће поглавље описује архитектуру коришћеног модела, процес тренирања и метрике за евалуацију перформанси генерисаног текста.

Четврто поглавље садржи детаље експерименталне поставке. Приказане су компоненте кодера и великог језичког модела, спојни модул (пројектор) и механизам маскирања пажње при прослеђивању информација. Пето поглавље посвећено је анализи постигнутих перформанси, где је дата квантитативна евалуација заснована на објективним метрикама и квалитативна анализа кроз конкретне примере генерисаних описа слика. Дискутоване су уочене предности и ограничења модела.

Шесто поглавље садржи изведене закључке и правце за даља истраживања.

4. Анализа са кључним резултатима

Мастер рад дипл. инж. Елхана Даздаревића се бави разматрањем мултимодалних MLLM/VLM модела заснованих на дубоком учењу (deep learning). Предмет мастер рада је генерисање описа слике применом AI модела. Циљ рада је имплементација визуелно-језичког приступа заснованог на ViT-GPT2 кодер-декодер архитектуре у сврху генерисања описа слике, уз употребу визуелног трансформера ViT (Visual Transformer) као визуелног кодера и генеративног трансформера GPT-2 (Generative Pre-Trained Transformer) као декодера. За поравнање визуелних карактеристика са векторским простором језичког модела коришћен је MLP (Multilayer Perceptron) пројектор. За реализацију мултимодалног модела је коришћен Python програмски језик уз употребу PyTorch библиотеке за дубоко учење. Модел је трениран и евалуиран помоћу

јавно доступног Flickr8k скупа података. Примењено је више објективних метрика (BLEU, METEOR, ROUGE-L, CIDEr) за квантитативну евалуацију.

Кључни доприноси у изради тезе су:

- 1) преглед и анализа постојећих решења у области мултимодалних и визуелно-језичких модела са фокусом на изазове визуелних халуцинација, сигурности, правичности и мултимодалног усклађивања,
- 2) развој и имплементација система за аутоматско генерисање описа слика заснованог на ViT-GPT2 кодер-декодер архитектури са MLP пројектором за поравнање репрезентација,
- 3) експериментална евалуација над скупом података уз примену објективних метрика у квантитативној анализи и поређење са резултатима из литературе,
- 4) спровођење квалитативне анализе и визуелизације пажње ради идентификације узрока халуцинација и анализе везе између визуелних сегмената и генерисаног текста.

5. Закључак и предлог

Кандидат Елхан Даздаревић је у свом мастер раду успешно реализовао експериментални софтверски систем за аутоматско генерисање текстуалних описа слика на основу њиховог визуелног садржаја, коришћењем комбинације визуелних и језичких модела. Добијени резултати указују на то да је за реалну процену квалитета модела неопходно упоређивати стандардне нумеричке показатеље са стварном семантичком тачношћу генерисаног текста, посебно због повремене појаве погрешних детаља у описима сложенијих сцена. Сprovedена тестирања су показала да предложени приступ, заснован на спајању визуелног кодера и језичког декодера преко пројекционог модула, има практичну примену у аутоматизацији индексирања и претраживања визуелног садржаја, као и у развоју мултимедијалних система за помоћ слабовидим особама. Кандидат Елхан Даздаревић је показао да може самостално да користи релевантну литературу, да препозна и дефинише проблематику, спроведе имплементацију и доноси селективне закључке.

На основу изложеног, Комисија за преглед и оцену мастер рада предлаже Комисији за студије II степена Електротехничког факултета у Београду да рад дипл. инж. Елхана Даздаревића прихвати као мастер рад и кандидату одобри јавну усмену одбрану.

Београд, 10.09.2026. године

Чланови комисије:

др Ана Гавровска Ванредни професор
сагласан, 09.09.2026.

др Горан Марковић Ванредни професор
сагласан, 10.09.2026.