



Veliki jezikovni modeli za digitalno humanistiko / Large language models for digital humanities: LLM4DH

Izroček projekta / Project Deliverable

Veliki jezikovni modeli za folkloristiko / Large language models for folkloristics

Št. Izročka / Deliverable No.	5.2.1
Datum oddaje / Delivery date	31. 3. 2026
Vrsta / Nature ¹ :	R
Nivo diseminacije / Dissemination level ¹ :	PU
Delovni Sklop / Work package	WP5 Izbrani izzivi digitalne humanistike / WP5 Challenges of digital humanities T5.2 Veliki jezikovni modeli za folkloristiko / Large language models for folkloristics
Avtor, Vodilni partner / Author, Lead partner	Polona Tratnik, IRRIS
Sodelujoči partnerji / Contributing partners	Polona Tratnik, IRRIS Darko Darovec, IRRIS Jan Babnik, IRRIS Jure Koražija, IRRIS Marjan Horvat, IRRIS Jasmina Rejec, IRRIS Žiga Oman, IRRIS Urška Lampe, IRRIS Anja Mlakar, IRRIS Matej Martinc, IJS Marko Robnik Šikonja, UL FRI Tjaša Arčon, UL FRI

¹Nature:

R = Report, P = Prototype, D = Demonstrator, O = Other (specify)

¹Dissemination level

PU = Public

CO = Confidential, only for members of the consortium (including ARIS)

Informacijska tabela projekta / Project information card	
Akronim projekta / Project acronym	LLM4DH
Naslov projekta / Project title	Veliki jezikovni modeli za digitalno humanistiko / Large language models for digital humanities
Spletna stran / Website	https://www.cjvt.si/llm4dh/
Št projekta / Project No.	GC-0002
Razpis / Call	Javni razpis za (so)financiranje Gravitacije, št: 5100-70/2024-6 z dne 7.5.2024 in Sprememb javnega razpisa za (so)financiranje Gravitacije, št. 5100-70/2024-11 z dne 29.5.2024.
Koodrinator projekta / Project coordinator	Marko Robnik-Šikonja
Trajanje projekta / Project duration	1.10.2025 – 30.9.2027, 36 mesecev / 36 months
Povzetek projekta / Abstract of the project	<p>Jezikovne tehnologije igrajo vse pomembnejšo vlogo na številnih področjih znanosti. Z bliskovitim napredkom velikih jezikovnih modelov se ti pozicionirajo kot osrednja tehnologija umetne inteligence z daljnosežnim vplivom ne le na znanstveno raziskovanje ampak tudi na širši družbeni razvoj in celotno družbo. Veliki jezikovni modeli, kljub svojim velikim zmožnostim, izkazujejo vrsto pomanjkljivosti, kot so nekonsistentni in napačni odgovori, velike računske zahteve, šibko poznavanje jezikov z manj viri, neprilagojenost za nekatere pomembne naloge in domene ter slabosti v razumevanju družbe, etike in človekovih potreb. Projekt bo tehnologijo velikih jezikovnih modelov izboljšal z različnimi znanji in podatki, jim zmanjšal računsko zahtevnost pri govornih tehnologijah in jim omogočil dostop do zanesljivih zunanjih informacij. Izboljšave v temeljnih zmožnostih velikih jezikovnih modelov bodo vodile do prebojnih dosežkov na več področjih digitalne humanistike: v jezikoslovju, leksikografiji, izobraževanju, analizi družbenih pojavov in sodobne zgodovine, političnih vedah, folkloristiki in pravu.</p> <p>Language technologies are increasingly important in many areas of science. With the rapid advances in large-scale language models, they are positioning themselves as a core AI technology with far-reaching impact on scientific research, wider societal development, and society. Despite their great capabilities, large language models exhibit several shortcomings, such as inconsistent and incorrect answers, high computational demands, weak knowledge of less-resourced languages, poor performance in some important tasks and domains, and weaknesses in understanding society, ethics, and human needs. The project will improve the technology of large language models by enriching them with diverse knowledge and data, using them to improve the performance of speech technologies, and giving them access to reliable external information. Improvements in the core capabilities of large language models will lead to breakthroughs in several areas of the digital humanities: linguistics, lexicography, education, analysis of social phenomena and contemporary history, political sciences, folkloristics, and law.</p>

Vodilni partner prprojekta/ Lead project partner	Univerza v Ljubljani, Fakulteta za računalništvo in informatiko (UL FRI):
Sodelujoči partnerji / Participating project partners	<p>Institut "Jožef Stefan" (IJS) Inštitut za novejšo zgodovino (INZ) Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja (IRRIS) Inštitut za kriminologijo pri Pravni Fakulteti v Ljubljani (IK) Univerza v Ljubljani, Filozofska fakulteta (UL FF) Univerza v Ljubljani, Fakulteta za elektrotehniko (UL FE) Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko (UM FERI)</p>
Opis aktivnosti 5.2.1 / Description of the Action 5.2.1	<p>Cilj naloge T5.2 je razviti metodologijo računalniške folkloristike, utemeljeno na velikih jezikovnih modelih (LLM), z namenom povečanja učinkovitosti in natančnosti primerjalnih raziskav ter omogočanja obsežnih diahronih, medjezikovnih in medkulturnih analiz. Ključni raziskovalni izziv predstavlja avtomatizacija analize pripovednih struktur pri preučevanju različnih različic zgodb skozi čas in v različnih kulturnih okoljih. Predvidena izročka sta nova metodologija za digitalno folkloristiko (M18) ter analize ritualov reševanja konfliktov (M36).</p> <p>The objective of Task T5.2 is to develop a computational folkloristics methodology based on large language models (LLMs), aimed at improving the efficiency and accuracy of comparative research and enabling large-scale diachronic, cross-linguistic, and cross-cultural analyses. The central research challenge is the automation of narrative structure analysis across different versions of stories over time and across cultures. The main deliverables include a novel methodology for digital folkloristics (M18) and analyses of conflict resolution rituals (M36).</p>

Zgodovina dokumenta / Document history			
Datum / Date	Verzija	Avtor / Author	Komentar / Comment
20. 3. 2026	1.0	Polona Tratnik	Izdelana verzija za končni pregled

Kazalo vsebine / Table of content

1. CILJI	5
2. REZULTATI	5
3. PUBLIKACIJE IN DISEMINACIJA	6

1. Cilji

Cilj naloge T5.2 je razviti metodologijo računalniške folkloristike, utemeljeno na velikih jezikovnih modelih (LLM), z namenom povečanja učinkovitosti in natančnosti primerjalnih raziskav ter omogočanja obsežnih diahronih, medjezikovnih in medkulturnih analiz. Ključni raziskovalni izziv predstavlja avtomatizacija analize pripovednih struktur pri preučevanju različnih različic zgodb skozi čas in v različnih kulturnih okoljih. Predvidena izročka sta nova metodologija za digitalno folkloristiko (M18) ter analize ritualov reševanja konfliktov (M36).

2. Rezultati

V okviru aktivnosti 5.2.1 smo uspešno razvili novo metodologijo za digitalno folkloristiko (M18), ki temelji na uporabi LLM in je strukturirana v treh fazah: (1) vzpostavljanje zanesljivih podatkovnih baz, (2) anotacija podatkov in metapodatkov ter (3) napredna analitična obdelava. V prvem letu projekta smo se osredotočili na prvo fazo in oblikovali štiri ključne digitalne korpuse: korpus etnografskih besedil, korpus slovenskih ljudskih pravljic iz zbirke ZRC SAZU (z izločitvijo problematičnih rokopisov), korpus revije Ciciban ter prilagojeni korpus Kmetijskih in rokodelskih novic (KRN).

Posebna pozornost je bila namenjena obravnavi heterogenega gradiva (tipkopisi, rokopisi, stare tiskovine), za kar smo razvili pristop, ki uporablja LLM za transkripcijo, normalizacijo in pripravo podatkov. Ta pristop omogoča nadaljnje obsežne diahrono, sinhrono in medkulturne analize folklornih pripovedi ter predstavlja pomemben metodološki napredek na področju digitalne humanistike.

Raziskovalne aktivnosti so bile podprte tudi z intenzivno diseminacijo rezultatov. Organizirana je bila mednarodna delavnica AI Methods for the Research of Folkloristic Narratives (12.–13. 6. 2025, UL FRI), na kateri je kot vabljeni predavatelj sodeloval Timothy Tangherlini (UC Berkeley). Rezultati raziskav so bili predstavljeni v številnih konferenčnih prispevkih, vabljenih predavanjih in znanstvenih publikacijah, vključno z objavljenimi članki, sprejetimi in oddanimi rokopisi, kar potrjuje znanstveno relevantnost in mednarodno odmevnost razvite metodologije.

3. Publikacije in diseminacija

V okviru aktivnosti 5.2 -veliki jezikovni modeli za folkloristiko smo opravili naslednje diseminacijske aktivnosti:

Organizacija mednarodne delavnice / Organization of international workshop: *AI Methods for the Research of Folkloristic Narratives*, 12. 6. 2025, FRI (IRRIS & FRI); vabljeni predavatelj / invited plenary speaker: Timothy Tangherlini (UC Berkeley).

Predstavitve in objavljeni povzetki s te delavnice / Workshop papers:

- ARČON, Tjaša, ROBNIK ŠIKONJA, Marko, TRATNIK, Polona. Motif detection using large language models : the Cinderella case study. V: TRATNIK, Polona (ur.). *AI methods for research of folkloristic narratives : international workshop : programme and book of abstracts* : [June 13, 2025, University of Ljubljana, Faculty of Computer and Information Science, & online (Zoom)]. Čentur: Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja: = Istituto IRRIS di ricerca, sviluppo, e strategie della societa, cultura e ambiente: = Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2025. Str. 7. Reflektor, Letn. 1 (2025), št. 2. ISBN 978-961-95505-4-0. https://www.iris.eu/wp-content/uploads/Reflektor_junij-2025_DIGITAL.pdf. [COBISS.SI-ID 242182915]
- BABNIK, Jan, MARTINC, Matej. Considering modes: semiotics and multimodal AI. V: TRATNIK, Polona (ur.). *AI Methods for Research of Folkloristic Narratives: international workshop: programme and book of abstracts: June 13, 2025, University of Ljubljana, Faculty of Computer and Information Science, & online (Zoom)*. Čentur: Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja: = Istituto IRRIS di ricerca, sviluppo, e strategie della societa, cultura e ambiente: = Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2025. Str. 9. Reflektor, letn. 1 (2025), št. 2. ISBN 978-961-95505-2-6. https://www.iris.eu/wp-content/uploads/Reflektor_junij-2025_DIGITAL.pdf. [COBISS.SI-ID 270916099]
- BABNIK, Jan, TRATNIK, Polona. The Dragon-slayer's narrative: structural kinship and discursive divergence. V: TRATNIK, Polona (ur.). *AI methods for research of folkloristic narratives : international workshop : programme and book of abstracts* : [June 13, 2025, University of Ljubljana, Faculty of Computer and Information Science, & online (Zoom)]. Čentur: Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja: = Istituto IRRIS di ricerca, sviluppo, e strategie della societa, cultura e ambiente: = Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2025. Str. 8. Reflektor, Letn. 1 (2025), št. 2. ISBN 978-961-95505-4-0. [COBISS.SI-ID 242779907]
- HORVAT, Marjan, KORAŽIJA, Jure, TRATNIK, Polona. Modeling deliberative values in narrative culture using LLMs. V: TRATNIK, Polona (ur.). *AI methods for research of folkloristic narratives: international workshop: programme and book of abstracts* : [June 13, 2025, University of Ljubljana, Faculty of Computer and Information Science, & online (Zoom)]. Čentur: Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja: = Istituto IRRIS di ricerca, sviluppo, e strategie della societa, cultura e ambiente: = Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2025. Str. 8. Reflektor, Letn. 1 (2025), št. 2. ISBN 978-961-95505-4-0. [COBISS.SI-ID 245423363]
- REJEC, Jasmina. Computational detection of moral values in Slovenian animal folktales using GPT-4 and lexicon-based methods. V: TRATNIK, Polona (ur.). *AI Methods for Research of Folkloristic Narratives: international workshop : programme and book of abstracts* : June 13, 2025, University of Ljubljana, Faculty of Computer and Information Science, & online (Zoom). Čentur: Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja: = Istituto IRRIS di ricerca, sviluppo, e strategie della societa, cultura e ambiente: = Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2025. Str. 8. Reflektor, letn. 1 (2025), št. 2. ISBN

978-961-95505-2-6. https://www.irris.eu/wp-content/uploads/Reflektor_junij-2025_DIGITAL.pdf. [COBISS.SI-ID 245162755]

- DAROVEC, Darko, ERGAVER, Angelika, KORAŽIJA, Jure, KOS, Veronika, OMAN, Žiga. The Earliest Accounts of the Americas and the Customary Conflict Resolution Analyzed by OpenAI and ChatGPT. V: TRATNIK, Polona (ur.). AI Methods for Research of Folkloristic Narratives: international workshop: programme and book of abstracts: June 13, 2025, University of Ljubljana, Faculty of Computer and Information Science & online (Zoom). Čentur: Inštitut IRRIS, 2025, str. 9. (Reflektor, letn. 1 (2025), št. 2). ISBN 978-961-95505-2-6.

Drugi konferenčni prispevki / Other conference papers:

- ARČON, Tjaša, ROBNIK ŠIKONJA, Marko, TRATNIK, Polona. Automatic detection of folkloristic motifs with large language models: the Cinderella tale. V: International conference AI for science 2025: Ljubljana, Slovenia, 22.09.2025-26.09.2025: including DS-2025, the SMASHing conference (4 tracks) and 4 additional tracks. Ljubljana: University of Ljubljana, Faculty of computer and information science, 2025. Str. 12-13. https://ai4science.si/wp-content/uploads/2025/09/AI4Science_Book_of_Abstracts-4.pdf. [COBISS.SI-ID 254172163]
- WILLAERT, Tom, OLBRICH, Eckehard, SANTAGIUSTINA, Carlo, ZECH, Nora, BABNIK, Jan, LIAO, Lingyu, ZHANG, Shuyu, LOUREIRO, Marina, LATIL, Brogan, STEWART, Michelle, et al. Affordances for expressing collective identities : the case of the 2024 French parliamentary elections. V: Visual methods for digital research : Digital Methods Summer School and Data Sprint 2024 : 1-12 July 2024, Amsterdam, Netherlands. Amsterdam: Digital Methods Initiative, 2024. 1 spletni vir, ilustr. <https://www.digitalmethods.net/Dmi/SummerSchool2024AffordancesCollectiveIdentities>. [COBISS.SI-ID 271076611]
- ALBRIGHT, Jonathan, AGGARWAL, Piyush, BABNIK, Jan, CATTANEO, Anna, HU, Haohan (Lily), LIAO, Lingyu, ZHU, Mengdi. SynthViz : enhancing interpretative summarization and visualization techniques. V: Visual methods for digital research: Digital Methods Summer School and Data Sprint 2024: 1-12 July 2024, Amsterdam, Netherlands. Amsterdam: Digital Methods Initiative, 2024. 1 spletni vir, ilustr. <https://www.digitalmethods.net/Dmi/Synthiz>. [COBISS.SI-ID 271113475]
- BABNIK, Jan. Structural semiotics at scale: computational approaches to online narrative analysis. V: LAMPE, Urška (ur.). Cultural memory, public spheres and social media: international workshop: programme and book of abstracts : [February 12-13, 2026, Hotel Grand Koper]. Elektronska izd. Čentur: Inštitut IRRIS za raziskave, razvoj in strategije družbe, kulture in okolja: = Istituto IRRIS di ricerca, sviluppo, e strategie della società, cultura e ambiente: = Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2026. Str. 15-16. Reflektor, letn. 2 (2026), št. 2. ISBN 978-961-95505-7-1. https://www.irris.eu/wp-content/uploads/Reflektor_02_2026_final1.pdf. [COBISS.SI-ID 270947331]
- BABNIK, Jan, TRATNIK, Polona. Affordances, modes, and modalities in shaping cultural memory on social media. V: The impact of social media on democracy: final conference of the EU project "Social Media for Democracy (SoMe4Dem) - Understanding the causal mechanisms of digital citizenship": Harnack-Haus, Berlin, 13. - 14. 1. 2026. Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.: München. 2026, 1 spletni vir. <https://www.mis.mpg.de/de/events/event/affordances-modes-and-modalities-in-shaping-cultural-memory-on-social-media>. [COBISS.SI-ID 266934275]
- MARTINC, Matej, BABNIK, Jan. Teaching vision-language models semiotics: toward a socio-semiotic framework for multimodal AI. V: International conference AI for science 2025: Ljubljana, Slovenia,

22. 09. 2025-26. 09. 2025: including DS-2025, the SMASHing conference (4 tracks), and 4 additional tracks. Ljubljana: University of Ljubljana, Faculty of computer and information science, 2025. Str. 13. https://ai4science.si/wp-content/uploads/2025/09/AI4Science_Book_of_Abstracts-4.pdf. [COBISS.SI-ID 251907075]

- BABNIK, Jan. Collective memory, narrative, structuralism, and generative AI: exploring the role of generative AI in narrative research. V: HORVAT, Marjan (ur.), ERGAYER, Angelika (ur.). International Scientific Workshop Cultural Memory in Online Public Discourse: between polarization and deliberation: 7–9 May 2025, Portorož, Slovenija. [S. l.]: Institute IRRIS for Research, Development and Strategies of Society, Culture and the Environment, 2025. Str. 10-11. <https://www.iris.eu/wp-content/uploads/SoMe4Dem-Pororoz-1F.pdf>. [COBISS.SI-ID 270928899]
- BABNIK, Jan. Deliberation, mediatization, and trust: notes toward a conceptual and methodological framework : lecture at the COST Action 22149, ChangeCode WG2 Meeting: Knowledge Exchange Event, 28-29 April, 2025 (Hybrid), Universitat de les Illes Balears, Mallorca, Spain. [COBISS.SI-ID 270923779]
- REJEC, Jasmina. Moral discourse and cultural memory in Slovenian animal tales within the Slavic narrative tradition. V: Fairy tales in Slavic countries : reception, production, circulation : international conference : Sapienza University of Rome, 11-13 February 2026 : book of abstracts. [Rome: s. n.], 2026. Str. 71. <https://talesrome2026.wordpress.com/abstracts/>. [COBISS.SI-ID 270893315]
- REJEC, Jasmina. Animal folk tales as vehicles of moral values and social critique. V: Humble theory and the power of the vernacular : 14th International Conference of Young Folklorists: 25th-26th September, 2025 University of Tartu : abstracts. Tartu: University of Tartu, 2025. Str. 37-38. ISBN 978-9985-4-1482-8. <https://sisu.ut.ee/wp-content/uploads/sites/867/Book-of-Abstracts.pdf>. [COBISS.SI-ID 270881027]
- TRATNIK, Polona: *Cinderella Between Worlds: Aesthetic Dialogues Across Chinese and Western Traditions*, International Symposium on Intercultural Dialogue and Contemporaneity of Aesthetics, Shenzhen University, Kitajska / China, 21.–24. 11. 2025

Vabljeno predavanje na tuji univerzi / Invited lecture at a foreign university:

- TRATNIK, Polona. Fairy Tales as Instruments of Social Formation: “Hansel and Gretel” Across Changing Social Orders. Beijing: Normal University Beijing, 3. december 2025. [Vabljeno predavanje].

Sprejeto v objavo / Accepted for publication:

- Arčon, Robnik Šikonja, Tratnik: *Large Language Models for Folktale Type Automation Based on Motifs: Cinderella Case Study, Fabula* (2026)

Oddano v objavo / Submitted:

- Machidon et al.: *Large Language Models for OCR in Cultural Heritage: A Comparative Study on Slovene Folkloristic Texts*

Objavljeni članki / Published articles:

- TRATNIK, Polona. Saint George, the Dragon Slayer: sacralized violence and the allegorical union of sacred and secular power. *Acta Histriae*. [Tiskana izd.]. 2025, letn. 33, št. 4, str.617-668, ilustr. ISSN 1318-0185. <https://ojs.zdjp.si/index.php/ah/article/view/275/155>, <https://zdjp.si/wp-content/uploads/2025/12/Acta-Histriae-33-2025-4-Polona-Tratnik-LOWRESS.pdf>, DOI: 10.19233/AH.2025.23. [COBISS.SI-ID 265969667], [Odprti dostop, JCR, SNIP, WoS, Scopus]
- TRATNIK, Polona. Flusser on artificial intelligence. *Flusser studies*. Nov. 2025, special twentieth anniversary iss., [no.] 40, str. 1-8. ISSN 1661-5719. <https://www.flusserstudies.net/sites/www.flusserstudies.net/files/media/attachments/tratnik.pdf>. [COBISS.SI-ID 265976835], [Odprti dostop]