

AI Methods for Research of Folkloristic Narratives

International workshop organized within projects Unraveling Sociocultural Dynamics through Computational Folkloristics (BI-US/24-26-034), Political Functions of Folktales (N6-0268), Artificial Intelligence for Digital Humanities (EU ERA Chair Grant no. 101186647 (AI4DH)), Large Language Models for Digital Humanities (GC-0002 (LLM4DH)), Practices of Conflict Resolution Between Customary and Statutory Law in the Area of Today's Slovenia and its Neighboring Lands (P6-0435), and Language Resources and Technologies for Slovene (P6-0411) by:

Institute IRRIS for Research, Development and Strategies of Society, Culture and Environment & University of Ljubljana, Faculty of Computer and Information Science (UL FRI)

Location: UL FRI, Lecture Hall P02

June 13, 2025, from 8:45 to 18:15

Zoom: <https://uni-lj-si.zoom.us/j/92230149490?pwd=gFEtauF3snPg9QeqgpxXiuPgywb5LF.1>



PROGRAM AND BOOK OF ABSTRACTS

Edited by Polona Tratnik

AI Methods for Research of Folkloristic Narratives

CONTENTS

CONCEPT	3
PROGRAM	5
Marko Robnik-Šikonja: Large Language Models for Analysis of Complex Phenomena.....	6
Tim Tangherlini: AI Archives: Leveraging Multilingual Large Language Models for Folklore Archives.....	7
Sándor Darányi, Thomas van Erven: An Integrated Text Processing Workflow for Folk Narrative Studies.....	8
Tjaša Arčon, Marko Robnik Šikonja, Polona Tratnik: Motif Detection Using Large Language Models: The Cinderella Case Study	9
Judith Veld: LLMs for Motif Classification in Folktales: A Prompt Engineering Approach	10
Jasmina Rejec: Computational Detection of Moral Values in Slovenian Animal Folktales Using GPT-4 and Lexicon-Based Methods	11
Marjan Horvat, Jure Koražija, Polona Tratnik: Modeling Deliberative Values in Narrative Culture Using LLMs....	12
Jan Babnik, Polona Tratnik: The Dragon-Slayer's Narrative: Structural Kinship and Discursive Divergence.....	13
Darko Darovec, Angelika Ergaver, Jure Koražija, Veronika Kos, Žiga Oman: The Earliest Accounts of the Americas and the Customary Conflict Resolution Analyzed by OpenAI and ChatGPT.....	14
Jan Babnik, Matej Martinc: Considering Modes: Semiotics and Multimodal AI.....	15
Antoine Doucet: Presentation of the Centre of Excellence in Artificial Intelligence for Digital Humanities	16
AUTHORS	17

AI Methods for Research of Folkloristic Narratives

CONCEPT

Institute IRRIS for Research, Development and Strategies of Society, Culture and Environment & University of Ljubljana, Faculty of Computer and Information Science

Workshop program and organizing committee: Polona Tratnik (chair), Marko Robnik-Šikonja, Darko Darovec.

In recent years, artificial intelligence has opened new pathways for interpreting large-scale cultural and historical data, enabling researchers to identify latent patterns, structures, and discursive shifts that would remain hidden through traditional methods. Folkloristic narratives—fairy tales, myths, legends, and oral stories—form one of the richest and most culturally embedded forms of human expression. Yet, despite their complexity and abundance, they have rarely been subjected to computational methods on a meaningful scale. Why should we analyze folktales with AI? Because these narratives carry encoded models of social behavior, political imagination, moral reasoning, and cultural identity—models that evolve over time and differ across societies. AI, particularly large language models (LLMs), now offers tools powerful enough to detect recurring motifs, classify tale types, and even extract embedded values from a diachronic and intercultural perspective.

The workshop *AI Methods for the Research of Folkloristic Narratives* brings together scholars from folklore studies, digital humanities, computer science, and cultural history to explore the potential of AI in the study of traditional narratives. It is organized within the framework of several research initiatives:

- the bilateral project Unraveling Sociocultural Dynamics through Computational Folkloristics (BI-US/24-26-034), conducted in cooperation with UC Berkeley,
- Artificial Intelligence for Digital Humanities (EU ERA Chair Grant no. 101186647 (AI4DH)),
- Political Functions of Folktales (ARIS N6-0268),
- Large Language Models for Digital Humanities (GC-0002 (LLM4DH)),
- Research programs Practices of Conflict Resolution Between Customary and Statutory Law in the Area of Today's Slovenia and its Neighboring Lands (P6-0435),
- and Language Resources and Technologies for Slovene (P6-0411).

The central aim is to develop interdisciplinary methodologies for analyzing folkloristic corpora using AI-based tools, combining both qualitative insight and quantitative rigor. The bilateral project specifically addresses the sociocultural dynamics behind the popularity and adaptation of specific folktales in different societies, including Slovenian narratives and their intercultural ties. It builds on the idea that tale structures, motifs, and character functions are not universal givens, but are culturally shaped responses to social, political, and historical conditions. Computational approaches—especially those using LLMs—enable large-scale diachronic and synchronic analysis of these narratives, helping us understand their transformations over time and across regions.

The workshop program includes three thematic panels:

- LLMs for folktale type index and motif detection, focusing on computational workflows and case studies such as Cinderella,
- LLMs for value extraction and deep structure analysis, targeting moral foundations, deliberative democratic values, and actantial/semiotic structures,
- and LLMs for studying complex cultural phenomena, including multimodal AI and conflict narratives.

Participants include leading scholars and pioneers in computational folkloristics, such as Prof. Timothy Tangherlini (UC Berkeley), Prof. Emer. Sándor Darányi (University of Borås), and researchers from the Institute IRRIS for Research, Development and Strategies of Society, Culture and Environment, University of Ljubljana, Faculty of Computer and Information Science and Faculty of Arts, University of Borås, Swedish School of Library and Information Science, University of Vienna, and the Jožef Stefan Institute. The workshop will conclude with a presentation of the Centre of Excellence in Artificial Intelligence for Digital Humanities by ERA Chair Prof. Antoine Doucet.

The event serves as both a knowledge-exchange platform and a foundation for building an international research network in computational folkloristics. It supports the European strategic priority of digital transformation and strengthens links between humanities, computer science, and AI research—demonstrating that modern technology can enrich our understanding of deep cultural memory.



AI4DH CENTRE OF EXCELLENCE IN AI
FOR DIGITAL HUMANITIES



AI Methods for Research of Folkloristic Narratives

PROGRAM

Time	Activity
08:45 - 09:00	Get together
09:00 - 09:45	WELCOME AND INTRODUCTION TO THE TOPIC
09:00 - 09:15	Welcome (Polona Tratnik)
09:15 - 09:45	Marko Robnik-Šikonja: Large Language Models for Analysis of Complex Phenomena
09:45 - 11:00	PLENARY TALK
09:45 - 10:45	Tim Tangherlini: AI Archives: Leveraging Multilingual LLMs for Folklore Archives
10:45 - 11:00	Discussion
11:00 - 11:30	Coffee break
11:30 - 12:15	PANEL 1: LLMS FOR FOLKTALE TYPE INDEX AND MOTIF DETECTION
11:30 - 12:00	Sándor Darányi, Thomas van Erven: An Integrated Text Processing Workflow for Folk Narrative Studies
12:00 - 12:30	Tjaša Arčon, Marko Robnik Šikonja, Polona Tratnik: Motif Detection Using LLMs: The Cinderella Case Study
12:00 - 12:30	Judith Veld: LLMs for Motif Classification in Folktales: A Prompt Engineering Approach
12:30 - 12:45	Discussion
12:45 - 14:00	Lunch break
14:00 - 15:45	PANEL 2: LLMS FOR VALUES AND DEEP STRUCTURE ANALYSIS
14:00 - 14:30	Jasmina Rejec: Computational Detection of Moral Values in Slovenian Animal Folktales Using GPT-4 and Lexicon-Based Methods
14:30 - 15:00	Marjan Horvat, Jure Koražija, Polona Tratnik: Modeling Deliberative Values in Narrative Culture Using Large Language Models
15:00 - 15:30	Jan Babnik, Polona Tratnik: The Dragon-Slayer's Narrative: Structural Kinship and Discursive Divergence
15:30 - 15:45	Discussion
15:45 - 16:15	Coffee break
16:15 - 18:00	PANEL 3: LLMS FOR STUDYING COMPLEX PHENOMENA
16:15 - 16:45	Darko Darovec, Angelika Ergaver, Jure Koražija, Veronika Kos, Žiga Oman: The Earliest Accounts of the Americas and the Customary Conflict Resolution Analyzed by OpenAI and ChatGPT
16:45 - 17:15	Jan Babnik, Matej Martinc: Considering Modes: Semiotics and Multimodal AI
17:15 - 17:45	PROJECT PRESENTATION by ERA Chair, Antoine Doucet: Centre of Excellence in Artificial Intelligence for Digital Humanities
17:45 - 18:00	Discussion
18:00 - 18:15	Wrap-up

Large Language Models for Analysis of Complex Phenomena

Currently, large language models (LLMs) are redefining methodological approaches in many scientific areas, including digital humanities. By digesting and analyzing large amounts of text, LLMs can provide insights into complex phenomena and questions. However, to get scientifically useful results, the studies involving LLMs have to be carefully methodologically designed and involve strict qualitative and quantitative evaluation. We will explain the working of LLMs needed to understand their performance. We will emphasize the need to establish trust in their performance when analyzing complex phenomena.

AI Archives: Leveraging Multilingual Large Language Models for Folklore Archives

In this brief presentation, I present some of the challenges we have confronted in several folklore archive digitization projects: ISEBEL (Intelligent Search Engine for Belief Legend), SAMLA (The Norwegian ethnographic archive digitization project), and the Berkeley Folklore Archive. I present the entire pipeline from physical archival materials through to broadly available search engines, and show how judicious use of multilingual LLMs can assist with many of the tasks. In particular, I show how an LLM can be used to help estimate the data schema for poorly labeled data, how these models can assist with both OCR correction and search over a corpus with multiple languages and dialects, how data embeddings can be used to develop a recommender / discovery system, and how network representations of the folklore data (e.g. fairy tales) can be "disentangled" to discover underlying latent meaning spaces in these complex collections.

An Integrated Text Processing Workflow for Folk Narrative Studies

To continue previous work, we introduce an open-source Python workflow which combines machine learning based text categorization with sentiment vs. social importance analysis, leading to information fusion. Its result is a 3-d content landscape instead of the typical 2-d content distribution maps computed by latent methods, having taken thereby a conceptual shift from vector spaces to vector fields whose semantic content can be coupled with graph representations for ontology building. In a recent experiment we tagged motifs by their computed average sentiment scores to construct a high-level landscape of the Thompson Motif Index (TMI), then visualized the narrative arcs of Aarne-Thompson-Uther (ATU) tale types as motif strings with specific 3-d coordinates on a test basis. Such narrative arcs can express e.g. sentiment intensity changes over plot progress by action types, a new kind of microscope for zooming in on new aspects of content or visualize plot unfolding as a sequence of type-specific motifs in TMI space for motif replacement/simplification testing. First experiments with the Ashliman's Folktexs / Annotated Folktales (AFT) dataset are ongoing.

Tjaša Arčon (University of Ljubljana, Faculty of Computer and Information Science) Marko Robnik Šikonja (University of Ljubljana, Faculty of Computer and Information Science), Polona Tratnik (Institute IRRIS & University of Ljubljana, Faculty of Computer and Information Science & Faculty of Arts)

Motif Detection Using Large Language Models: The Cinderella Case Study

Artificial intelligence approaches are being intensively adapted to many research areas, including digital humanities. We built a methodology for modern, effective, and large-scale analyses in folkloristics. Using several modern machine learning and natural language processing approaches, we automatically detected motifs in a large collection of Cinderella variants and analyzed their similarities and differences with the help of clustering and dimensionality reduction. The results show that large language models can detect complex interactions in tales, enabling computational analysis of significantly larger text collections than previously possible. Our approach opens a path to large-scale analyses of folktales and fairy tales, their cross-lingual comparisons, and the generalization of motifs beyond folkloristics.

LLMs for Motif Classification in Folktales: A Prompt Engineering Approach

This project explored the use of publicly available Large Language Models (LLMs) for classifying folktale motifs into a more generalized hierarchy. I developed the hierarchy based on the ATU index, WordNet, and LLM input, then evaluated motif extraction by ChatGPT and DeepSeek using zero-, one-, few-shot, and code-style prompts. Results showed moderate accuracy and challenges with abstraction, hierarchical depth, and gender or cultural bias. The study underscores the limitations of current classification systems and highlights the need for closer collaboration between expert annotators and LLM researchers.

Computational Detection of Moral Values in Slovenian Animal Folktales Using GPT-4 and Lexicon-Based Methods

This pilot study presents a computational approach to identifying moral values in approximately 60 Slovenian animal folktales, using GPT-4 zero-shot classification and a lexicon-based method grounded in Moral Foundations Theory. The tales were translated into English via ChatGPT to enable keyword-based lexicon matching, while GPT-4 was applied directly to the original Slovenian texts. Outputs were interpreted using chain-of-thought reasoning. A preliminary expert review was conducted to assess the alignment between machine-generated and human-assigned moral labels. The results were compiled into a unified dataset and analyzed across twelve moral dimensions. This exploratory analysis offers initial insight into the feasibility of combining large language models and lexicon methods for the study of moral content in folktales.

Marjan Horvat (Institute IRRIS), Jure Koražija (Institute IRRIS), Polona Tratnik (Institute IRRIS & University of Ljubljana, Faculty of Computer and Information Science & Faculty of Arts)

Modeling Deliberative Values in Narrative Culture Using LLMs

This paper applies large language models (LLMs) to analyze comparatively Slovenian versions of three folk tales - *The Giant Turnip*, *Mojca Pokrajculja*, and *The Ant and the Grasshopper*. Through a structured prompt based on deliberative democracy standards, we identify value-laden cognitive matrices shaping inclusion, mutual respect, responsibility, and common good. Combining narrative roles and discursive moral logics, the method reveals diachronic shifts and cultural specificities in collective decision-making. Our preliminary results prove LLMs' effectiveness for comparative cultural analysis, enabling the reconstruction of proto-political imaginaries embedded in traditional storytelling and their relevance for contemporary democratic practice.

The Dragon-Slayer's Narrative: Structural Kinship and Discursive Divergence

This presentation explores dragon-slayer narratives that share a similar internal structure yet differ discursively. By comparing legends, folktales, chivalric romance, and novel, it investigates how these structurally related tales have different discursive functions, shaped by their narrative logic, ideological framing, and cultural context across religious, folkloric, and popular domains. It combines Greimas's semiotic square with the actantial scheme in LLM-assisted analysis of dragon-slaying tales. By *Overlaps Mapping* it tracks 15 structural relations between six actantial roles to detect dominant and co-occurring overlap patterns across different dragon-slaying tales; through *Semiotic Square Application* it maps the semiotic square on core narrative oppositions (e.g., True Faith vs. False Belief); and across all actantial roles to uncover internal contrasts and meta-structural patterns.

The Earliest Accounts of the Americas and the Customary Conflict Resolution Analyzed by OpenAI and ChatGPT

This preliminary study proposes that vengeance is a universal, ritualized form of conflict resolution leading to peace, rather than a violent trait of so-called primitive cultures. Drawing on premodern European and colonial American case studies, our investigation compares customary rituals—such as banishment and asylum—across societies. Using OpenAI’s ChatGPT, early sources by Sahagún, Durán, and de las Casas were analyzed structurally through Greimas’s framework, supplemented by Propp, Lévi-Strauss, and the moral codes of outlaw heroes (Hobsbawm, Seal). The glossary from *The Language of Vengeance* (Darovec, Ergaver & Oman, 2017) provided a key tool for interpreting historical narratives of conflict and peace.

Considering Modes: Semiotics and Multimodal AI

This presentation examines Generative AI and multimodal models through the lens of socio-semiotic multimodality and its potential applications. By contrasting machine learning paradigms of *medium* with social semiotic theories of *mode* we offer a preliminary assessment of how GAI engages with complex meaning-making practices and inter-semiotic translation. In doing so, we raise critical questions concerning the tension between qualitative and quantitative approaches, while also exploring avenues for methodological refinement that might reconcile these not-so-dissimilar paradigms. The presentation will focus on the development of a vision language model (VLM) for Slovenian, how to train the model to answer questions about visual content, and how to reduce hallucination and improve reasoning capabilities of the model in a low resource environment with scarce data availability.

Antoine Doucet (University of Ljubljana, Faculty of Computer and Information Science & University of La Rochelle)

Presentation of the Centre of Excellence in Artificial Intelligence for Digital Humanities

The ERA Chair project AI4DH is establishing an interdisciplinary European Centre of Excellence for Artificial Intelligence in Digital Humanities. The Centre will provide humanities researchers with support in using AI tools, machine learning, and data analysis. It will offer state-of-the-art infrastructure for developing AI models, including large language models, and expert assistance in creating new approaches. The project will also address the ethical use of AI and foster inclusive and culturally diverse digital environments. The Centre will also support project applications in the field of artificial intelligence for digital humanities and will enable applicants to visit the University of Ljubljana.

AI Methods for Research of Folkloristic Narratives

AUTHORS

Tjaša Arčon is a researcher at the Lab for Machine Learning and Language Technologies at the Faculty of Computer and Information Science, University of Ljubljana. Her background is in linguistics, language teaching, and cognitive science, and I currently explore the application of large language models (LLMs) in the humanities.

Jan Babnik works as a researcher, editor, and educator. He is a researcher at Institute IRRIS, editor-in-chief of *Membrana Journal*, and a director of the Membrana Institute. He is primarily concerned with visual-discursive theories and research practices, multimodality, and the philosophy of visual culture.

Prof. Emer. Sándor Darányi (MSc in Agriculture, MA in Library and Information Science, PhD in Computational Ethnography) has been employed in the academe since 1985, worked in science diplomacy during the 6th EU Research Framework Programme, initiated and/or has contributed to several national and international RTD projects during the 7th EU Research Framework Programme, Horizon 2020, and Horizon Europe. His research interests include digital humanities and folk narrative studies.

Darko Darovec is a Full Professor of History at the University of Maribor, Senior Research Fellow, and Director of the IRRIS Institute. He has decades of experience in leading national and international research and innovation projects, such as PI of research program ARIS P6-0435, Action Chair of COST CHANGECODE (2023–2027), INTERREG, Marie Skłodowska-Curie Fellow at Ca' Foscari University in Venice. He authors or (co-)authors 17 scientific monographs, more than 100 original scientific articles and book chapters.

Antoine Doucet, the ERA Chair Holder at the University of Ljubljana since 2025, is a Professor of Computer Science at the University of La Rochelle, France, since 2014, he was Vice-Rector for European University and led a research team of about 50 people. PI of numerous research projects, including H2020 NewsEye, a research group leader in La Rochelle Université, as well as director of research and education department in ICT at the University of Science and Technology of Hanoi (a.k.a., Vietnam-France University).

Angelika Ergaver is a research associate with the Institute IRRIS. Her work focuses on mechanisms of dispute settlement in the South Adriatic region, with particular emphasis on legal customs of reconciliation among Montenegrin and Northern Albanian kinship communities during the early modern period.

Thomas van Erven is an independent IT contractor. He has worked for research institutes and universities in the Netherlands, Sweden, Belgium and the US on EU, national, and international projects. His professional interests and background include automation, deployment, data pipelines in combination with enrichment tools, and AI.

Marjan Horvat (PhD) is a sociologist and research associate at the IRRIS Institute. He leads the IRRIS research team in the Horizon Europe project *Social Media for Democracy (SoMe4Dem)* and serves as a member of the Management Committee and scientific representative of IRRIS (Grant Holder Institution) in the *COST Action CHANGECODE – Research Network for Interdisciplinary Studies of Transhistorical Deliberative Democracy*.

Jure Koražija is a Junior Researcher at the IRRIS Institute and a doctoral student in sociology. His dissertation, under the mentorship of Dr. Marjan Horvat, investigates transhistorical deliberative democracy using artificial intelligence tools and contemporary computational analytics.

Veronika Kos is a junior researcher at Institute IRRIS and a doctoral student in history at the Faculty of Arts, University of Maribor. Supervised by Professor Dr Darko Darovec, she explores the role of women in dispute settlement in the Slovene lands during the early modern period.

Matej Martinc is a postdoctoral researcher at the Jožef Stefan Institute, specializing in the interdisciplinary areas of Natural Language Processing (NLP), Digital Humanities, and Computer Vision. He is currently developing AI systems that can analyze and generate human language. He also works on integration of visual and textual information.

Žiga Oman is a research associate with the Institute IRRIS and assistant professor at the University of Maribor, Faculty of Arts. His research focuses on early modern social history of Inner Austria (1500–1800), particularly on dispute settlement and control of violence, including their semantic field. In 2025 he has been awarded Marie Skłodowska-Curie Action which he will carry out in 2025-27 at the University of York (UK).

Jasmina Rejec is a doctoral student at the University of Ljubljana and a Junior Researcher at the Institute IRRIS. Her interdisciplinary research on Slovenian animal folk tales combines computational methods with qualitative sociocultural analysis to explore moral values, narrative motifs, and their cultural significance.

Marko Robnik-Šikonja is a Professor of Computer Science and Informatics at the University of Ljubljana, Faculty of Computer and Information Science, and head of Machine Learning and Language Technology Lab. He is (co)author of more than 250 scientific publications (with more than 9,500 citations), and recipient of several rewards for his work. He has contributed to and led several national and EU projects and authored several data mining software packages and language resources.

Timothy R. Tangherlini is the Elizabeth H. and Eugene A. Shurtleff Chair in Undergraduate Education at UC Berkeley, where he is Distinguished Professor in the Dept of Scandinavian and the School of Information. He is also the faculty archivist for the Folklore Program. His publications include *Interpreting Legend: Danish Storytellers and their Repertoires* and *Danish Folktales, Legends and Other Stories*. An early contributor to the field of computational folkloristics, his current work focuses on rumor, legend, belief and conspiracy theories.

Polona Tratnik is Senior Research Fellow and Head of the Centre for Digital Humanities at the Institute IRRIS and a Full Professor at the Faculty of Arts at the University of Ljubljana. She was a Fulbright Visiting Scholar and has led several research projects on social and political functions of fairy tales, folktales, and arts. She is an author or co-author of 11 scientific books (plus 2 are forthcoming), including *In the Land of Fairy Tales: On the Social Functions of Fairy Tale Narratives* (2024), as well as approx. 100 scientific articles and book chapters.

Judith Veld is a student of Cognitive Science at the University of Vienna but currently writing her master's thesis at the University of Ljubljana. She has a background in Psychology and gender studies, with research interest spanning folklore, climate change, psychopathology, and neurophenomenology.