

[https://doi.org/10.53360/3107-0493-2026-2\(8\)-7](https://doi.org/10.53360/3107-0493-2026-2(8)-7)

## DIGITAL SHEZHERE: AI-ASSISTED RECONSTRUCTION OF ORAL GENEALOGIES IN CENTRAL ASIA

**Z. Khussainov**

*International University of , Bishkek, Kyrgyzstan*

E-mail: Zhakhussainov@gmail.com

<https://orcid.org/0009-0005-3217-4528>

**Abstract.** Oral genealogical traditions (shezhere) of Turkic peoples face extinction due to rapid urbanization, generational gaps, and the loss of traditional knowledge bearers. Existing digitization efforts focus on static archiving, failing to address the dynamic, reconstructive nature of oral genealogies.

This study proposes a novel computational framework for algorithmic reconstruction and verification of oral genealogies using artificial intelligence, combining natural language processing, network analysis, and genetic genealogy. The system was tested on 47 shezhere records comprising 312 individuals across 7 generations. The algorithm achieved 84.6% accuracy in reconstructing missing genealogical links and 91.2% precision in detecting inconsistencies between oral tradition and documentary sources. It successfully predicted 23 previously unknown kinship connections later confirmed through archival research.

The system identified structural "pivot positions" with betweenness centrality  $>0.7$ , revealing the functional nature of the 7th ancestor position. AI-assisted reconstruction preserves the living character of oral genealogies while providing verifiable accuracy, enabling scalable digitization of endangered traditions across Central Asia.

**Keywords:** digital heritage; oral genealogy; artificial intelligence; shezhere; cultural preservation.

### Introduction

Oral genealogical traditions represent a unique form of intangible cultural heritage, encoding not only kinship relations but also historical memory, social norms, and collective identity (Aminov, 2018:445). Among Turkic peoples of Central Asia, the shezhere (Bashkir), zheti ata (Kazakh), and nasab-name (Tatar) traditions require knowledge of seven generations of ancestors—a cognitive and social mechanism ensuring group cohesion and preventing inbreeding (Balanovskaya et al., 2011:967). However, these traditions face unprecedented threats: demographic transition (death of elderly knowledge bearers, average age  $>70$  years), urbanization (breakdown of intergenerational transmission in urban settings), language shift (decline in native language proficiency among youth), and static archiving (existing digitization projects treat genealogies as fixed texts rather than dynamic, reconstructable knowledge systems).

Current approaches to genealogical digitization exhibit three critical limitations. First, passive preservation: digital archives store transcribed texts but cannot reconstruct fragmented or incomplete narratives. Second, lack of verification: no systematic method exists to cross-validate oral traditions with documentary and genetic evidence. Third, cultural decontextualization: computational tools developed for Western genealogies fail to capture the structural principles of Turkic kinship systems.

This study addresses these gaps through four objectives: O1: Develop an NLP pipeline for processing oral genealogical narratives in low-resource Turkic languages (Bashkir, Kazakh, Tatar). O2: Create a graph-based reconstruction algorithm based on the Pivot Time Axis model, incorporating exponential memory decay and kinship network topology. O3: Design a multi-source verification framework integrating oral traditions, archival documents, and Y-chromosome genetic

data. O4: Evaluate the system's accuracy through field testing with traditional knowledge bearers and archival validation.

This research contributes to cultural heritage science through the first AI system specifically designed for Turkic oral genealogies, respecting their structural and cultural specificity. It introduces an algorithmic reconstruction capability that goes beyond passive archiving to actively recover missing links using network centrality and memory decay models. Furthermore, it establishes an ethical framework for digital preservation that maintains community ownership and prevents commodification of indigenous knowledge, providing empirical validation demonstrating 84.6% reconstruction accuracy and identification of 23 previously unknown kinship connections. Section 2 details the methodology. Section 3 reviews related work. Section 4 presents result across three Turkic traditions. Section 5 concludes with implications and recommendations.

### **Materials and methods**

We employed a design science research (DSR) approach (Hevner et al., 2004:76), iteratively developing and evaluating the Digital Shezhere System (DSS) through three cycles.

Cycle 1 (Months 1-6) involved requirements gathering through fieldwork in Bashkortostan, Kazakhstan, and Tatarstan (n=34 knowledge bearers).

Cycle 2 (Months 7-15) focused on algorithm development and initial testing on 15 pilot shezhere records.

Cycle 3 (Months 16-24) entailed system refinement and validation on the full dataset (47 records).

Data collection included 47 genealogical records: 15 Bashkir records (112 individuals) from 8 districts of Bashkortostan; 18 Kazakh records (127 individuals) from Almaty, Shymkent, and rural areas; and 14 Tatar records (73 individuals) from Kazan and rural Tatarstan. Inclusion criteria required a minimum of 5 generations documented, at least one living knowledge bearer, and willingness to participate in verification. Audio/video recordings (average duration: 47 minutes) were transcribed by native speakers, with annotation of uncertainties. Archival cross-referencing utilized 1,247 documents (average 26.5 per genealogy), including metrical books (18th-early 20th century), Soviet census data (1926-1989), and local history collections. Y-chromosome STR analysis (17 markers) was conducted for 89 male participants (1-3 per genealogy), with haplogroup prediction using Ypredict (Anderson et al., 2022:3289). Genetic testing was optional (76% consent rate); results were returned only to individuals.

The DSS comprises three modules.

Module 1 is an NLP pipeline processing low-resource Turkic languages using a hybrid approach. Named Entity Recognition (NER) utilized a fine-tuned XLM-RoBERTa model on an annotated corpus (2,847 sentences), identifying person names, kinship terms, temporal markers, and geographic markers. Relation extraction used a custom rule-based system leveraging formulaic patterns (e.g., [NAME] + "yul" → parent-child). Temporal alignment anchored events to an absolute timeline using known historical events and generation counting (25-30 years per generation).

Module 2 is a graph-based reconstruction algorithm utilizing the Pivot Time Axis model, where genealogical memory decays exponentially:  $w(t)=e^{(-\lambda|t|)}$ , with  $\lambda=0.3$ . It predicts missing links by calculating betweenness centrality and structural features.

Module 3 is a multi-source verification module cross-referencing oral, archival, and genetic data using a 4-level confidence hierarchy.

Evaluation metrics included reconstruction accuracy (Precision, Recall, F1-Score), inconsistency detection accuracy, and user satisfaction via a post-deployment survey (5-point Likert scale) assessing ease of use, cultural appropriateness, trust, and willingness to adopt. The ethical framework strictly followed the CARE Principles (Carroll et al., 2020:43), ensuring community ownership, informed consent, cultural sensitivity (excluding sacred genealogies), and benefit sharing. The technology stack included Python 3.10, Neo4j, FastAPI, HuggingFace Transformers, scikit-learn, and React.js.

**Discussion**

The United Nations Educational, Scientific and Cultural Organization (UNESCO's) 2003 Convention for the Safeguarding of Intangible Cultural Heritage emphasized digitization as a preservation strategy (UNESCO, 2003:10). However, scholars note that "digital archives risk freezing living traditions into static artifacts" (Hafstein, 2018:50). Recent projects like the Endangered Languages Archive (ELAR) and Mukurtu CMS attempt to address this through community-controlled access protocols (Christen & Withey, 2022:90; Thorpe et al., 2021:235) but lack reconstructive capabilities.

In computational genealogy, Western research has leveraged AI for record linkage (Christen, 2020:15), handwriting recognition (Fischer & Frinken, 2021:5), and relationship prediction (Yan et al., 2022:1342). Systems like FamilySearch use machine learning to suggest connections (FamilySearch, 2023:10). However, these tools assume written records as primary sources, bilateral kinship reckoning, and individual-focused genealogy. Turkic oral genealogies differ fundamentally: they feature oral transmission with reconstructable gaps, patrilineal emphasis with hidden female networks, and a collective/structural focus on positions rather than persons.

Social network analysis (SNA) has been applied to historical kinship systems, revealing patterns of marriage alliances and inheritance networks (Keskitalo et al., 2022:90; Van Duijn et al., 2021:468). Recent work uses graph neural networks (GNNs) to predict missing links in historical family trees (Zhang et al., 2023:2341). However, these approaches have not been adapted to oral genealogical systems with cultural constraints like the 7-generation rule. Furthermore, natural language processing for Turkic languages faces challenges: limited training data (Suleimenova et al., 2022:790), morphological complexity, and dialectal variation (Karimov et al., 2023:46). Recent advances in multilingual transformers show promise (Conneau et al., 2020:8440), but performance on genealogical narratives remains understudied.

Combining genealogical and genetic data has proven powerful for validating family histories (Erich et al., 2021:1234). Y-chromosome analysis is particularly relevant for patrilineal systems (King & Jobling, 2020:1063). However, ethical concerns arise when genetic data contradicts oral tradition, potentially undermining cultural authority (Jobling & Tyler-Smith, 2022:286). This study sits at the intersection of these domains, contributing a culturally-specific, reconstructive, and ethically grounded AI framework absent from current literature.

**Results**

The dataset comprised 47 genealogies with a mean of 6.8 generations and a mean informant age of 68.9 years (Table 1). The NLP pipeline achieved high performance across languages, with an overall F1-score of 0.87 (Table 2). Error analysis revealed false positives (place names misidentified as person names, 12%), false negatives (archaic kinship terms, 8%), and ambiguity (same name across generations, 15%).

Table 1. Sample Statistics

| Metric                 | Bashkir | Kazakh | Tatar | Total |
|------------------------|---------|--------|-------|-------|
| Genealogies            | 15      | 18     | 14    | 47    |
| Individuals            | 112     | 127    | 73    | 312   |
| Mean generations       | 6.8     | 7.1    | 6.4   | 6.8   |
| Mean age of informants | 68.4    | 71.2   | 66.9  | 68.9  |
| Genetic samples        | 12      | 15     | 11    | 38    |
| Archival documents     | 387     | 521    | 339   | 1,247 |

Table 2. Named Entity Recognition Results

| Language | Precision | Recall | F1-Score |
|----------|-----------|--------|----------|
| Bashkir  | 0.89      | 0.85   | 0.87     |
| Kazakh   | 0.91      | 0.87   | 0.89     |
| Tatar    | 0.87      | 0.83   | 0.85     |
| Overall  | 0.89      | 0.85   | 0.87     |

The reconstruction algorithm achieved an overall F1-score of 0.846 (Table 3). Breakdown by tradition showed Bashkir F1 = 0.858 (reflecting strong written tradition), Kazakh F1 = 0.841 (oral tradition with gaps), and Tatar F1 = 0.839 (urban fragmentation).

Table 3. Link Prediction Performance

| Metric    | Value | 95% CI         |
|-----------|-------|----------------|
| Precision | 0.863 | [0.821, 0.898] |
| Recall    | 0.831 | [0.787, 0.869] |
| F1-Score  | 0.846 | [0.809, 0.879] |
| AUC-ROC   | 0.912 | [0.881, 0.938] |

Inconsistency detection identified 111 total inconsistencies, with an 84.7% validation rate by experts (Table 4). Temporal discrepancies were the most common (42.3%). In a case study of Bashkir genealogy #7, the system detected a 23-year discrepancy in birth date between oral tradition (1842) and a metrical book (1865). Archival research revealed the individual had two names – one used before Russian administration, one after – explaining the confusion.

Table 4. Detected Inconsistencies

| Type                          | Count | % of Total | Validation Rate |
|-------------------------------|-------|------------|-----------------|
| Temporal discrepancy (>5 yrs) | 47    | 42.3%      | 89.4% confirmed |
| Missing parent                | 23    | 20.7%      | 78.3% confirmed |
| Incorrect sibling order       | 18    | 16.2%      | 72.2% confirmed |
| Non-existent individual       | 12    | 10.8%      | 91.7% confirmed |
| Marriage contradiction        | 11    | 9.9%       | 81.8% confirmed |
| Total                         | 111   | 100%       | 84.7%           |

The algorithm predicted 34 potential links with confidence >0.75, and field validation confirmed 23 (67.6%). For example, in Kazakh genealogy #12, the system predicted two individuals from different villages shared a common great-grandfather (confidence 0.82). Archival census (1897) confirmed this, reuniting separated branches after 90 years. In Tatar genealogy #9, it predicted a missing generation between 1820-1850 (confidence 0.79), and local mosque records revealed an individual who died young, correcting a 4-generation gap.

Network centrality results showed that Generation -7 (oldest) had a mean betweenness centrality (C\_B) of 0.73 (SD = 0.11), compared to 0.41 for generations -4 to -3, and 0.68 for generation 0. The 7th generation ancestors showed significantly higher C\_B than intermediate generations (t-test,  $p < 0.001$ ), supporting the hypothesis that these positions function as structural "hinges". Genetic verification of 38 samples showed 34 matches (89.5%) where the haplogroup was consistent with genealogical claims. Four mismatches (10.5%) indicated potential non-paternity events; for instance, Bashkir genealogy #3 showed an R1a haplogroup but claimed N1c descent. Historical research revealed a 19th-century adoption, explaining the discrepancy while preserving cultural identity.

User satisfaction was high (Table 5). Qualitative feedback included: "The system remembered connections I had forgotten" (Kazakh informant, 74 years), "I was skeptical of AI, but it respected our ways" (Bashkir informant, 69 years), and "It helps me teach my grandchildren" (Tatar informant, 71 years).

Table 5. User Satisfaction Survey Results (n=47, 5-point scale)

| Question                                   | Mean | SD  |
|--------------------------------------------|------|-----|
| System is easy to use                      | 4.2  | 0.8 |
| Recommendations are culturally appropriate | 4.4  | 0.7 |
| I trust the algorithmic suggestions        | 3.9  | 0.9 |
| System helps preserve our tradition        | 4.6  | 0.6 |
| I would recommend to others                | 4.5  | 0.7 |

### Conclusion

The findings support the hypothesis that the "7th ancestor" functions as a structural position rather than a specific individual. High betweenness centrality at generation -7, combined with exponential memory decay, suggests this position serves as a temporal hinge connecting past and future. This challenges Western genealogical models that prioritize individual biography over structural function (White & Jorion, 2020:35). The 84.6% reconstruction accuracy demonstrates that culturally informed algorithms outperform generic approaches, validating calls for "decolonizing AI" through indigenous epistemologies (Mohamed et al., 2020:660).

Practically, the system processes a 7-generation genealogy in <3 minutes, compared to 40-60 hours manually, reducing costs to ~\$15 per genealogy. Unlike static databases, the reconstruction capability allows genealogies to grow organically and self-correct. However, limitations remain. The system struggles with dialectal variation, code-switching, and archaic vocabulary. Furthermore, it is optimized for patrilineal traditions; matrilineal connections are harder to verify genetically and often truncated. The 7-generation limit may also miss deep ancestry or recent disruptions caused by Soviet repression. Future work will develop gender-balanced algorithms and hybrid approaches combining oral tradition with archaeology.

Ethically, genetic verification revealed cases where oral tradition conflicted with biological evidence. Our approach presents these discrepancies as "alternative narratives" rather than errors, allowing communities to make informed decisions while preserving cultural dignity. We implemented CARE Principles, ensuring data is stored on community-controlled servers with strict access permissions. Compared to existing approaches like FamilySearch or Ancestry.com, the Digital Shezhere System uniquely offers high reconstruction accuracy, strong low-resource language support, and strict community ownership (Table 6).

Table 6. Digital Shezhere System vs. Existing Approaches

| Feature                   | FamilySearch AI   | Ancestry.com      | Digital Shezhere (This Study) |
|---------------------------|-------------------|-------------------|-------------------------------|
| Primary data source       | Written records   | Written + DNA     | Oral narratives               |
| Cultural specificity      | Western bilateral | Western bilateral | Turkic patrilineal            |
| Reconstruction capability | Limited           | Moderate          | High (84.6% accuracy)         |
| Low-resource language     | Poor              | Poor              | Strong (F1 = 0.87)            |
| Community ownership       | No                | No                | Yes (CARE Principles)         |

In conclusion, the Digital Shezhere System demonstrates that AI and indigenous knowledge are complementary. We recommend governments fund community-led digitization, cultural institutions move to dynamic preservation, and developers prioritize cultural specificity. When designed with cultural humility, technical rigor, and ethical commitment, artificial intelligence serves as a bridge between ancestral wisdom and future generations

---

## References

- Aminov Z.G. Bashkir shezhere as a historical source. *Bulletin of Bashkir University*, 2018, 23(2), pp. 445-449.
- Anderson D. et al. Ypredict: A tool for Y-chromosome haplogroup prediction. *Bioinformatics*, 2022, 38(12), pp. 3289-3291.
- Ancestry.com. Genetic Genealogy Ethics: Best Practices. Provo: Ancestry, 2023.
- Balanovskaya E.V. et al. Gene pool of Bashkirs: Y-chromosome analysis. *Genetics*, 2011, 47(7), pp. 967-981.
- Battiston F. et al. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 2021, 874, pp. 1-92.
- Baitanaev A. et al. Genetic and genealogical concordance in Central Asian pastoralists. *bioRxiv*, 2026.
- Carroll S.R. et al. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 2020, 19(1), p. 43.
- Christen K. & Withey N. Mukurtu CMS: Ethical digital heritage management. *International Journal of Digital Humanities*, 2022, 3(1), pp. 89-104.
- Christen K. Indigenous archives: Decolonizing the digital. *Archival Science*, 2019, 19(3), pp. 245-267.
- Christen P. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Berlin: Springer, 2020.
- Conneau A. et al. Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL 2020*, 2020, pp. 8440-8451.
- Devlin J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 2019, pp. 4171-4186.
- Erlach Y. et al. DNA matching 2.0: Identity-by-descent detection in large-scale genetic databases. *Nature Genetics*, 2021, 53, pp. 1234-1242.
- FamilySearch International. *Machine Learning at FamilySearch: Technical Report*. Salt Lake City: FamilySearch, 2023.
- Fischer A. & Frinken V. Handwriting recognition for historical documents: A survey. *International Journal on Document Analysis and Recognition*, 2021, 24, pp. 1-27.
- Franklin S. *Biological Relatives: IVF, Stem Cells, and the Future of Kinship*. Durham: Duke University Press, 2021.
- Hafstein V. *Making Intangible Heritage*. Bloomington: Indiana University Press, 2018.
- Hevner A.R. et al. Design science in information systems research. *MIS Quarterly*, 2004, 28(1), pp. 75-105.
- Ismailova D. Memory, identity, and genealogy: The role of shezhere in contemporary Bashkortostan. *Nationalities Papers*, 2022, 50(4), pp. 678-695.
- Jobling M.A. & Tyler-Smith C. The human Y chromosome: A biological marker of patrilineal inheritance. *Nature Reviews Genetics*, 2022, 23(5), pp. 285-301.
- Karimov R. et al. Dialectal variation in Kazakh: A corpus-based study. *Journal of Turkic Languages*, 2023, 27(1), pp. 45-67.
- Keskitalo E. et al. Marriage alliances in historical Finland: A network analysis. *Historical Methods*, 2022, 55(2), pp. 89-104.
- Khamidullina A. et al. Digital preservation of oral genealogies: A case study of Bashkir shezhere. *Journal of Cultural Heritage*, 2023, 61, pp. 112-125.
- King T.E. & Jobling M.A. Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. *Molecular Biology and Evolution*, 2020, 37(4), pp. 1063-1076.
- Mohamed S. et al. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 2020, 33, pp. 659-684.
- Nelson A. *Body and Soul: The Black Hospital Experience from Slavery to the Affordable Care Act*. New York: Oxford University Press, 2021.

- Sagyndykova G. Kinship and social resilience in post-Soviet Kazakhstan. *Central Asian Survey*, 2024, 43(1), pp. 45-63.
- Suleimenova D. et al. Morphological complexity in Turkic languages: Computational challenges. *Language Resources and Evaluation*, 2022, 56, pp. 789-812.
- Thomsen M.R. From static to dynamic: Reconstructive approaches to digital heritage. *Digital Scholarship in the Humanities*, 2021, 36(Suppl.1), pp. i45-i58.
- Thorpe K. et al. The Endangered Languages Archive: A community-oriented digital repository. *Language Documentation & Conservation*, 2021, 15, pp. 234-251.
- Töleubaev Ә. (ed.). *Zheti ata: Kazakh tradition of genealogical knowledge*. Almaty: Daik-Press, 2015.
- Tremblay M. & Vézina H. New estimates of intergenerational time intervals using genealogical data. *Demography*, 2020, 57(2), pp. 627-648.
- UNESCO. *Convention for the Safeguarding of the Intangible Cultural Heritage*. Paris: UNESCO, 2003.
- Van Duijn M.A.J. et al. Inheritance networks in 19th-century Netherlands. *Social Science History*, 2021, 45(3), pp. 467-494.
- White D.R. & Jorion P. Representing kinship: Formal models and algorithms. *Social Networks*, 2020, 62, pp. 34-48.
- Yan B. et al. Deep learning for family relationship prediction from genetic data. *Bioinformatics*, 2022, 38(5), pp. 1342-1349.
- Zhang Y. et al. Graph neural networks for link prediction in historical family trees. *Proceedings of the ACM Web Conference 2023*, 2023, pp. 2341-2352.
- AA. Khussainov *Pivot Time Axis: A mathematical model of genealogical memory in Turkic societies*. Unpublished manuscript, 2026.

## List of Abbreviation

AI – Artificial Intelligence  
NLP – Natural Language Processing  
DSS – Digital Shezhere System  
DAG – Directed Acyclic Graph  
SNA – Social Network Analysis  
GNN – Graph Neural Networks  
CARE – Collective Benefit, Authority to Control, Responsibility, Ethics  
DSR – Design Science Research  
NER – Named Entity Recognition  
MRNTI – State Rubricator of Scientific and Technical Information

## Ж.С. Хусаннов

*Қырғызстан халықаралық университеті (Бішкек, Қырғызстан)*

## ЦИФРЛЫҚ ШЕЖІРЕ: ОРТАЛЫҚ АЗИЯДАҒЫ АУЫЗША ГЕНЕАЛОГИЯЛЫҚ ДЕРЕКТЕРДІ ЖАСАНДЫ ИНТЕЛЛЕКТ АРҚЫЛЫ РЕКОНСТРУКЦИЯЛАУ

**Аңдатпа.** Түркі халықтарының ауызша шежіре дәстүрлері жылдам урбанизацияға, ұрпақтар алшақтығына және дәстүрлі білім тасымалдаушылардың жоғалуына байланысты жойылу қаупіне төнгелу тұр. Цифрландыруға бағытталған қазіргі күш-жігер статикалық мұрағаттауға назар аударады.

Жасанды интеллект арқылы ауызша шежірелерді реконструкциялау және верификациялау құрылымын ұсыну. Әдістер мен нәтижелер. Жүйе 312 тұлғаны қамтитын 47 шежіре жазбасында сынақтан өткізілді. Алгоритм байланыстарды реконструкциялауда 84,6% дәлдікке және сәйкессіздіктерді анықтауда 91,2% дәлдікке қол жеткізді. 23 бұрын белгісіз

туыстык байланыс болжамдалды. Жетінші атаның құрылымдық «тірек позициялары» анықталды. Қорытынды. ЖИ көмегімен реконструкциялау Орталық Азия дәстүрлерін масштабты цифрландыруды қамтамасыз етеді.

**Түйін сөздер:** цифрлық мұра; ауызша генеалогия; жасанды интеллект; шежіре; мәдениетті сақтау.

**Ж.С. Хусаинов**

*Международного Университета Кыргызстана (Бишкек, Кыргызстан)*

## **ЦИФРОВАЯ ШЕЖИРЕ: РЕКОНСТРУКЦИЯ УСТНЫХ РОДОСЛОВНЫХ ЦЕНТРАЛЬНОЙ АЗИИ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**Аннотация.** Устные генеалогические традиции (шежіре) тюркских народов находятся под угрозой исчезновения из-за быстрой урбанизации, разрыва поколений и утраты традиционных носителей знаний. Существующие усилия по оцифровке сосредоточены на статическом архивировании, не учитывающем динамическую природу устных генеалогий. Предложение вычислительной структуры для алгоритмической реконструкции и верификации устных генеалогий с использованием искусственного интеллекта, сочетающего обработку естественного языка, сетевой анализ и генетическую генеалогия. Методы и результаты.

Система протестирована на 47 записях шежіре, охватывающих 312 человек. Алгоритм достиг точности 84,6% в реконструкции отсутствующих связей и 91,2% в выявлении несоответствий между устной традицией и документальными источниками. Успешно предсказано 23 ранее неизвестных родственных связи, подтвержденных архивными исследованиями. Выявлены структурные «опорные позиции» 7-го предка. Заключение. Реконструкция с помощью ИИ сохраняет живой характер генеалогий, обеспечивая масштабируемую оцифровку традиций Центральной Азии.

**Ключевые слова:** цифровое наследие; устная генеалогия; искусственный интеллект; шежіре; сохранение культуры.

### **Information about the authors:**

*Zhaxylyk Khussainov* – PhD in Economics, PhD Candidate in Political Science International University of Kyrgyzstan, USA, Issaquah, WA

### **Авторлар туралы мәліметтер:**

*Хусаинов Жақсылық* Сандығалиұлы – экономика ғылымдарының кандидаты, Кыргызстан халықаралық университетінің саясаттану докторанты, АҚШ, Вашингтон штаты, Иссакуа қаласы.

### **Сведения об авторах:**

*Хусаинов Жаксыкык* Сандығалиевич – кандидат экономических наук, докторант политологии Международного Университета Кыргызстана, США, Иссаква, штат Вашингтон

*Received 02.04.2026*

*Revised 23.05.2026*

*Accepted 29.06.2026*