ZEYREK
Tarih Araştırmaları Dergisi

AI and The Future of the Translational Technologies: Tools for the Historians

Yapay Zekâ ve Çeviri Teknolojilerinin Geleceği: Tarihçinin Takım Çantası

Harun TUNCER*

Abstract

This article explores the intersection of artificial intelligence and Ottoman studies, focusing on how AI-assisted technologies transform the ways historians engage with historical texts. Following the alphabet reform of the early Turkish Republic, Ottoman Turkish materials—including manuscripts, printed works, and archival records—became largely inaccessible to modern readers, resulting in a profound rupture between scholars and their textual heritage. The study argues that the challenge lies not merely in the alphabet change but also in the reliability of critical editions and the identification of authentic manuscripts. After tracing the evolution of textual criticism in Turkey since the Second Constitutional Period, the paper examines how recent digital platforms have begun to bridge this historical disconnection. Through a comparative analysis of four major systems—Transleyt, Transkribus, Vesiqari, and AKİS—it assesses

Araştırma makalesi/
Research Article

Geliş tarihi/*Date of Arrival:*
19 Nisan/April 2025

Kabul tarihi/*Date of Acceptance:*
3 Kasım/November 2025

* Doç. Dr., İstanbul Üniversitesi, Türkiyat Araştırmaları Enstitüsü, Türk Tarihi, İstanbul/ Türkiye

ORCID ID: 0000-0001-8810-0310
E-posta/E-mail: harunt@istanbul.edu.tr

Atıf/*Citation*: Harun Tuncer, "AI and The Future of the Translational Technologies: Tools for the Historians" *Zeyrek Tarih Araştırmaları Dergisi*, 3 (Aralık 2025), s. 3-20.
10.5281/zenodo.18004495

their technical infrastructure, accuracy, and usability in processing Ottoman Turkish texts. The discussion extends to the ethical and methodological implications of relying on opaque "black box" algorithms, emphasizing that while AI can accelerate transcription, transliteration, and data extraction, it cannot replace human interpretation or contextual judgment. Ultimately, these technologies should be viewed not as substitutes for scholarly expertise but as complementary tools that expand the historian's capacity to read, analyze, and reinterpret the documentary legacy of the Ottoman world.

Keywords: Ottoman Turkish, artificial intelligence, digital humanities, critical edition, historiography

ÖZ

Bu makale, yapay zekâ teknolojileri ile Osmanlı araştırmaları arasındaki kesişim alanını ele alarak, özellikle tarihçilerin tarihî metinlerle kurduğu ilişkiyi dönüştüren dijital imkânları inceliyor. Türkiye Cumhuriyeti'nin ilk yıllarında gerçekleştirilen Harf Devrimi sonrasında Osmanlı Türkçesiyle kaleme alınmış el yazmaları, basılı eserler ve arşiv belgeleri modern araştırmacılar için büyük ölçüde erişilemez hale geldi; böylece bilim insanları ile metinler arasındaki bağ büyük oranda zayıfladı. Çalışma, sorunun yalnızca alfabe değişikliğinden ibaret olmadığını; güvenilir tenkitli neşirlerin hazırlanması ve sahih nüshaların tespiti gibi meselelerin de aynı derecede önem taşıdığını vurguluyor. II. Meşrutiyet döneminden itibaren Türkiye'de metin neşri geleneğine de temas eden makale, son yıllarda bu tarihî kopukluğu gidermeye çalışan dijital platformların imkânlarını değerlendiriyor. Transleyt, Transkribus, Vesiqari ve AKİS gibi dört sistemin/projenin teknik altyapılarını, doğruluk oranlarını ve Osmanlı Türkçesi metinleri işleme kapasitelerini karşılaştırmalı biçimde incelemeye çalışıyor. Ayrıca, bu sistemlerin "kara kutu" niteliğindeki algoritmalarına duyulan güven sorunu ile etik ve metodolojik boyutları da tartışıyor. Neticede, yapay zekâ teknolojilerinin tarihçinin yerini alacak araçlar değil; aksine, tarihçinin okuma, çözümleme ve yorumlama becerisini genişleten tamamlayıcı araçlar olduğu sonucuna varıyor.

Anahtar Kelimeler: Osmanlı Türkçesi, yapay zekâ, dijital beşerî bilimler, tenkitli neşir, tarih yazımı.

1. Introduction

Texts written in Ottoman Turkish have long constituted indispensable sources for historians. Over time, however, these texts have become inaccessible to a broad readership—not only due to the language itself, but also owing to the writing systems and terminology used. The alphabet reform initiated in the early years of the Turkish Republic, along with the efforts to simplify the language, aimed to facilitate literacy. Yet these changes also weakened the connection with the linguistic and cultural heritage of the past. Geoffrey Lewis famously described this rupture as a "tragic success," arguing that the transition to the Latin script severed ties with a vast corpus of cultural memory.

[1] This disconnection went beyond script reform: the vocabulary was narrowed, syntactic structures were broken down, and, in the end, even educated individuals were no longer able to comprehend routine documents written just a century ago.[2] As a result, documents written in Ottoman Turkish became enigmatic to many scholars, even in the humanities.

However, the main difficulty with Ottoman texts lies not merely in the change of alphabet but also in the reliability of their publication[3]—that is, in the ability to identify and reproduce the *original* or the most accurate surviving copy of a work. As textual critics have noted, "the first task in editing a text is to identify all available manuscripts and then determine which of them will be used for the edition."[4] In this respect, catalogues such as TÜYATOK and those prepared by various institutions or individuals must be carefully

1 Geoffrey Lewis, *Trajik Başarı: Türk Dil Reformu*, tr. by M. Fatih Uslu, (İstanbul: Çevribilim Yayınları), 2019, p. 9.

2 Demircioğlu, Cemal. *Çeviribilimde Tarih ve Tarihyazımı*, (İstanbul: Boğaziçi University Press), 2016; Karaca, Taha Niyazi. *Tarihçi Geçmişi Kurgulamak: Kuram, Tasarım, İnşa*, (İstanbul: Kronik Kitap), 2020

3 The practice of publishing *critical editions* in Turkey began during the Second Constitutional Era (1908-1918), when significant efforts were made that rivaled contemporary European scholarship. For a detailed discussion of the development of textual criticism and edition practices in Turkey, see Ahmet Emin Saraç, "Modern Metin Tenkidi ve Tesisi Usulünün Türkiye'ye Girişi", *Türk Dili ve Edebiyatı Dergisi*, v. 63, no. 2, 2023, pp. 93-113.

4 M. Fatih Köksal, "Metin Neşrinin Ana Esasları", *TÜBAR*, XXXI (2012 Spring), p. 182.

examined, and while digital access has made this task considerably easier, electronic means alone should never suffice. Once the manuscripts are identified, all extant copies should, as far as possible, be obtained; next, they must be classified into *manuscript families*—often through the criterion of "shared errors"—and one representative or *lead copy* should be chosen from each group. Moreover, the ongoing methodological debate on *critical editions* must be acknowledged: some editors seek to produce a "sound text" with minimal annotation, whereas others regard it as essential to document textual variants, clarify obscure passages, and provide critical commentary.[5] These approaches are not mutually exclusive but rather complementary. Indeed, considering the frequent scribal errors, omissions, and inconsistencies in punctuation or vocalization, publishing a text without such critical comparison is academically precarious. Therefore, the ultimate aim of digital tools should extend beyond mere transcription to include manuscript collation, variant analysis, and the automatic generation of *stemma codicum* (manuscript genealogies).

Historians have always been occupied with texts. Yet when manuscripts or printed materials were not published or transcribed into the Latin script, accessing, reading, and analyzing these sources demanded an extraordinary amount of time and effort. For a long time, this labor remained the exclusive domain of experts, while the number of interested readers remained limited. Consequently, historical texts came to be treated merely as archival material rather than as living sources that could engage wider audiences. However, such texts do not only hold relevance for historical inquiry; they also provide vital clues for fields such as literature, economics, law, medicine, and social life.[6]

Recent advancements in digital technology have partially alleviated these difficulties. Systems capable of recognizing, interpreting, and even translating texts have been developed. In particular,

5 Beşşâr Avvâd Mârûf, "Tenkitli Metin Neşrinin Esasları", çev. Ahmet Eşer, *BAİD* (Balıkesir İlahiyat Dergisi) 14 (Aralık 2021), pp. 562-563.

6 Sadık Yazar, "Bakir Bir Araştırma Sahası Olarak Osmanlı Tercüme Geleneği." *Türk Dili ve Edebiyatı Dergisi*, 60/1 (2020), pp. 153-178.

AI-based tools have begun to assume roles traditionally performed by human experts in transcription and interpretation. Although these tools cannot replace the historian's intellectual judgment, they have eased the researcher's burden, drawn attention to overlooked materials, and opened new paths for analysis.[7]

Digital platforms developed within this context have redefined the historian's engagement with old-script texts. Text recognition, script conversion, keyword tracking, and the extraction of names and locations are now tasks that can be facilitated by software. Yet these conveniences have raised new questions: To what extent are these tools reliable? How successful are they in processing Ottoman Turkish, with its intricate linguistic structure? And are their outputs suitable for scholarly use?

This study seeks to answer these questions. It undertakes a comparative evaluation of four AI-powered platforms—Transkribus, Transleyt, Vesiqari, and AKİS—that have been developed for the analysis of Ottoman Turkish texts. By identifying each platform's capabilities, limitations, and potential areas of improvement, this study aims to illuminate how digital tools are reshaping the landscape of historical research in the age of artificial intelligence.

2. Challenges in Working with Historical Texts

The primary material of the historian is the text, which often carries the weight of archival documentation. Over time, however, such texts have posed challenges to historians themselves. Documents written in Ottoman Turkish—such as official records, imperial decrees, newspapers, letters, memoirs, and scholarly treatises—proved difficult to comprehend not only because of their script but also due to their complex syntactic structures, conceptual vocabulary, and semantic density. These difficulties became even more pronounced

7 Hughes-Warrington, Marnie, and Jo Guldi. "The Historian's Role in an Age of AI." Network in Canadian History & Environment, 2024. https://niche-canada.org/2024/11/21/the-historians-role-in-an-age-of-ai/ (Erişim: 07.04.2025); Spina Salvatore, "Artificial Intelligence in archival and historical scholarship workflow: HTS and ChatGPT," https://arxiv.org/abs/2308.02044v1 (Eririm: 14.04.2025).

when one considers the variety of scripts used and the evolutionary nature of the language itself.[8]

In manuscript texts, the challenges were even more acute than in printed works. Scripts such as *rik'a*, *ta'līq*, or *divānī* often featured overlapping letters, irregular line spacing, and ink smudges, all of which impeded legibility. Abbreviations, regional pronunciations transcribed phonetically, and idiomatic expressions with shifting meanings further complicated the task of deciphering authorial intent.

The issue was not confined to paleography. Following the alphabet reform, newer generations were no longer trained to read texts written in the Arabic script. This was not an incidental outcome but rather a deliberate policy decision with far-reaching consequences: it led not only to unfamiliarity with the script but also to alienation from the older form of the language. Indeed, the language simplification movement -actively promoted by state institutions beginning in the 1930s- resulted in the systematic pruning of the lexicon. In a short period, many people lost the ability to understand not only classical texts but even newspaper articles printed in the 1940s.[9]

This rupture has affected not only historians but also researchers working in literature, sociology, law, and even the natural sciences. Texts written in Ottoman Turkish preserve not just stories of the past but also its scientific reasoning, conceptual frameworks, administrative structures, and socio-cultural fabric. The inability to access such texts has meant a detachment not only from understanding the past but also from rethinking and reconstructing it.[10]

Digitization initiatives have offered a partial solution. Libraries, archives, and research institutions have begun digitizing their collections. Yet if these documents are made available solely as "digital

8 Esma F. Bilgin Taşdemir, Zeynep Tandoğan et. al. "Automatic Transcription of Ottoman Documents Using Deep Learning." In: Sfikas, G., Retsinas, G. (eds) *Document Analysis Systems. DAS 2024. Lecture Notes in Computer Science*, vol 14994. Springer, Cham, 2024, pp. 2-3.

9 Lewis, *Trajik Başarı*, pp. 9-12.

10 Fatma Aladağ and Elif Derin Can, "Programlamadan Yapay Zekâya Osmanlı Türkçesi için Metin Tanıma ve Otomatik Transkripsiyon," *Dijital Beşerî Bilimler ve Osmanlı Çalışmaları*, ed. Yunus Uğur, (İstanbul: Vakıfbank Kültür Yayınları), 2023, p. 71.

images," the researcher's engagement with the text remains superficial. For an image to be transformed into knowledge, it must either be transcribed manually or processed through optical character recognition (OCR) and translation tools.[11]

However, the distance between a digital image and reliable knowledge is often greater than it appears. Digitization is merely a starting point; the real challenge lies in converting digitized content into meaningful and trustworthy information. This can be achieved either through the labor of trained individuals or with the aid of AI-driven software systems. At this critical juncture, platforms such as Transleyt, Transkribus, Vesiqari, and AKİS have emerged -each with distinct capacities for making historical materials intelligible through computational means. However, it seems essential for the historian to be equipped with knowledge of philology and palaeography.[12]

3. Artificial Intelligence Technologies and Historical Research

Historiography has long enjoyed a rich tradition in both method and practice. Yet this tradition has increasingly found itself reshaped by the emergence of new technological capabilities. In particular, digitization efforts have transformed not only how historical documents are accessed, but also how they are interpreted, contextualized, and ultimately understood. Initially perceived as a matter of convenience, this transformation has gradually led to fundamental shifts in research methods. The historian's role is no longer confined to uncovering traces of the past—it now also involves reassembling and reinterpreting these traces using the tools of the present.[13]

In this evolving landscape, artificial intelligence (AI)–based applications have begun to find a place in historical research. These tools, capable of processing large datasets, classifying texts, detecting

11 Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage, 2014, p. 7.

12 Taha Niyazi Karaca, *Tarihçi Geçmişi Kurgulamak: Kuram, Tasarım, İnşa*, (İstanbul: Kronik Kitap), 2020, pp. 245-246.

13 Adam Crymble, *Technology and the Historian: Transformations in the Digital Age*, (Urbana: University of Illinois Press, 2021), p. 2; David M. Berry, *The Philosophy of Software: Code and Mediation in the Digital Age*, (New York: Palgrave Macmillan, 2011), p. 118.

patterns, and, to a certain extent, interpreting meaning, have opened new paths of inquiry. Subfields such as natural language processing (NLP), machine learning, and deep learning have all contributed to redefining the relationship between the historian and historical texts.

3.1. Natural Language Processing (NLP)

NLP refers to the attempt to teach machines how to "understand" human language. While it was initially applied to contemporary languages, NLP has also found applications in the processing of historical and even extinct languages. In a language layered as Ottoman Turkish, identifying word structures, idioms, synonyms, and period-specific usages was once a labor-intensive task limited to the expertise of lexicons and individual scholars. NLP systems have significantly eased this burden by accelerating transcription, enabling semantic analysis, and revealing lexical networks.[14]

However, using NLP on historical texts brings its own challenges. Non-standardized spelling, inconsistent punctuation, and missing diacritics often hinder the machine's ability to extract meaning from context. Moreover, identifying idiomatic expressions or rhetorical strategies unique to a particular era requires more than technical intervention; it necessitates human discernment and interpretive sensitivity. Thus, while NLP offers considerable assistance, it ultimately defers interpretive authority to the historian.[15]

3.2. Machine Learning and Deep Learning

Machine learning enables algorithms to detect patterns from data without being explicitly programmed for each case. Deep learning, in turn, allows for the layered and sophisticated analysis of these patterns using neural networks. These techniques have proven particularly useful in areas such as handwritten text recognition (HTR),

14 Michael Piotrowski, *Natural Language Processing for Historical Texts*. San Rafael, CA: Morgan & Claypool, 2012, p. 1.

15 Peter A. Stokes, "Digital Approaches to Paleography and Book History: Some Challenges, Present and Future," *Frontiers in Digital Humanities* 2/5 (2015), pp. 1-3, https://doi.org/10.3389/fdigh.2015.00005.

document classification, topic modeling, and the identification of personal and geographic names. Tasks that once required historians to comb through thousands of pages manually can now be completed within minutes.

Among such tools, Transkribus has marked a significant step forward in the automatic processing of manuscripts. By allowing users to train custom models for specific scripts, Transkribus has not only provided a ready-to-use system but has also fostered an interactive environment for scholarly collaboration. Meanwhile, platforms such as Transleyt have opened new avenues for historians by offering Latin transliteration and semantic analysis for texts written in the Arabic script.

The use of these technologies has altered the way historical knowledge is constructed. Documents that were previously deemed unreadable to the untrained eye have become machine-readable. Yet this development has raised an important question: To what extent do machine-generated outputs align with historical truth? The answer lies not in technological sophistication alone, but in the critical judgment and interpretive authority of the scholar.[16]

4. A Comparative Analysis of Four Platforms for the Transliteration and Transleation of Ottoman Turkish Texts

The digitization and processing of historical texts is not merely a technical operation; it also constitutes a scholarly and cultural undertaking. The software platforms developed for this purpose are not just tools -they reflect particular epistemological assumptions and play an active role in shaping how historical knowledge is produced. This section offers a comparative analysis of four platforms currently used to process texts written in Ottoman Turkish: Transleyt, Transkribus, Vesiqari, and AKİS. Each is evaluated in terms of its script recognition performance, transcription accuracy, user interface, and its potential contribution to historical research.

16 Bahaeddin Eravcı, H. Mustafa Eravcı, Efe Can Tanrısever, "Tarih Araştırmaları ve Yapay Zeka," *AUSBD*, 2024; "Yapay Zeka ve Sosyal Bilimler Öğretim" Özel Sayısı, pp. 10-11.

4.1. Transleyt: Machine-Assisted Reading of Printed Texts

Transleyt emerged as a multilingual AI platform, iniciated just following the ChapGpt, capable of transliterating, analyzing, and even simplifying Ottoman Turkish texts, as its predecessor.[17] It performed well with printed texts, particularly those typeset in *naskh* script. In practice, Transleyt would often present not only a Latin-script transliteration of a given text but, in some cases, also a modernized version –if the user requests. This feature, while helpful for some, raised concerns for scholars who required access to the original phrasing for purposes of interpretation and critical analysis.

Despite this, Transleyt's speed and accessibility made it a valuable tool for researchers engaging with historical variants of Turkish. In one example, a keyword search was conducted across Ottoman newspapers published between 1890 and 1915 using terms such as "plague," "cholera," and "quarantine." Transleyt quickly extracted relevant passages and offered a usable Latin-script output. Although expert supervision remained necessary, the platform proved highly effective for initial data collection and thematic mapping.

4.2. Transkribus: Model-Based Approaches for Manuscripts

Transkribus, a European initiative, was designed specifically for the automatic transcription of manuscripts. It distinguished itself by allowing users to train custom recognition models tailored to particular scripts and handwriting styles. As such, it offered not only a ready-made solution but also an interactive learning environment.

Transkribus, an AI-powered tool, facilitates historical document processing through handwritten text recognition (HTR), automatic transcription, and thematic annotation. The platform allows users to transcribe handwritten and printed texts, train custom models for different documents, and tag both structural and semantic features. It also supports output in formats such as TEI, TXT, PDF, and Word, and is accessible through a web-based interface. In 2023, the

17 Olgun Gündüz, "Edebiyat Sosyolojisi Değişmelerinde Yaklaşım Değişimleri." In *Cumhuriyetin Yüzüncü Yılında Sosyal Bilimler: Süreklilik ve Dönüşüm Sempozyumu Bildiri Kitabı*, ed.Bilal Güzel. Ankara Sosyal Bilimler Üniversitesi Yayınları, 2024, p. 410.

first open-access HTR model for printed Ottoman Turkish texts was developed by Süphan Kırmızıaltın, Fatma Aladağ, and Elif Derin and made available via Transkribus. Further details on this model and related projects can be found on the Digital Ottoman Corpora website. New users can register for free and begin with 500 credits to upload documents and perform automated transcriptions. After transcription, users may manually correct the output and apply thematic annotations. Though its user interface requires a degree of technical proficiency, Transkribus remains a valuable long-term investment for scholars working with handwritten sources. Researchers can use it to create annotated, searchable corpora of their own materials.[18]

4.3. Vesiqari: Limited to Print, Defeated by *Rik'a*

Vesiqari, as its name implies, was built as a document-centered recognition platform. Its capabilities were limited to printed texts in *naskh* script. The system accepted image uploads and returned a Latin-script output; however, this output was typically restricted to letter-by-letter transliteration, with no semantic processing or contextual refinement.

In manuscripts written in *rik'a*, the platform performed very poorly. It failed to distinguish individual characters and could not maintain proper line integrity. In some instances, over 70% of the output consisted of unintelligible character strings.

This outcome revealed that the platform remained in its early stages of development and lacked the comprehensive modeling required for diverse scripts. Its closed training infrastructure, limited dataset, and lack of user customization severely restricted its usability.

4.4. AKİS: A Collaborative Project with Academic Aspirations

AKİS project is a collaborative research initiative hosted by Sabancı University's Data Analytics Research and Application Center (VERİM) and Digital Humanities Lab (DHLab), with contributions

18 Elif Derin, "TRANSKRIBUS ile Osmanlı Türkçesi'nin Dijitalleştirilmesi", https://www.digitalottomanstudies.com/post/kopyası-transkribus-ile-osmanlı-türçesi-nin-dijitalleştirilmesi?lang=tr&utm_source=chatgpt.com (Erişim: 10.04.2025)

from Istanbul Medeniyet University and the University of Vienna. The project aims to develop an artificial intelligence–based system capable of automatically transcribing Ottoman Turkish texts—both handwritten and printed in the Arabic script—into the Latin alphabet. This effort addresses a fundamental challenge in the humanities: the manual transcription of Ottoman texts, a laborious process currently limited to a small group of specialists.

AKİS seeks to combine deep-learning techniques with large, custom-built datasets to recognize Ottoman handwriting (in *nesih*, *rik'a*, and other styles) and printed typefaces with high accuracy. The project involves multiple stages, including image preprocessing, line segmentation, labeling, model training, and the development of an accessible web interface resembling tools such as Google Translate. Through this platform, researchers will be able to upload document images—such as newspapers or manuscripts—and receive automatic Latin-script transcriptions.

Once completed, AKİS will make pre-1928 Ottoman Turkish materials accessible to a much wider audience, significantly reducing the time and effort required for research in history, literature, art, architecture, political science, and sociology. By bridging traditional philological expertise with machine learning, the project aims to democratize access to cultural heritage and create lasting impact across both academic and educational domains.

5. Ethical and Methodological Considerations

AI-assisted text processing systems have undoubtedly lightened the historian's workload. However, they have also introduced a set of intellectual and methodological challenges. First among these is the question of trust: How reliable are the outputs generated by such systems for scholarly research? Most of these platforms function as so-called black boxes—they do not disclose the internal logic of their decision-making processes to the user, thereby limiting transparency.[19]

19 Yiliyaer Abudureyimu, Yücel Oğurlu, "Yapay zekâ uygulamalarının kişisel verilerin korumasına dair doğurabileceği sorunlar ve çözüm önerileri," *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 20/41 (2021), 768-771.

This lack of transparency becomes particularly problematic in tasks such as translation and transcription, where the historian's engagement with the text is direct and interpretive. In a complex and historically rich language like Ottoman Turkish, any decontextualized output—whether a mistranslation or an unrequested simplification—can distort historical meaning. As observed in several cases, some platforms did not merely transliterate the text; they also "modernized" it by altering vocabulary and orthography. While these interventions may enhance accessibility for general readers, they compromise the authenticity of the source material for academic purposes.

Secondly, the role of the historian is increasingly brought into question. While these tools offer automated or semi-autonomous solutions, they also risk marginalizing the researcher's interpretive engagement. Yet history is not merely the accumulation of facts; it is also a process of intellectual confrontation with the text. Algorithms cannot perceive nuance; they are blind to intertextual references, rhetorical devices, and conceptual subtlety. Interpretation, in its fullest sense, remains the prerogative of the human intellect. Far from rendering the historian obsolete, these technologies have made their judgment even more essential.

A third concern involves accessibility. Some platforms are commercial; others are restricted to partner institutions. This creates a digital divide, particularly affecting researchers in developing regions. Yet history is a shared heritage, and access to historical materials ought to be open and equitable. In this light, the development of open-source systems and publicly funded projects appears not only desirable but necessary.[20]

6. Conclusion

Working with Ottoman Turkish texts demands both technical proficiency and intellectual labor. These texts are not merely relics of the past—they contain a conceptual richness capable of shaping

20 Kitchen, *The Data Revolution*, 39.

contemporary thought. However, the alphabet reform, language simplification policies, and modern educational paradigms have weakened scholars' access to this corpus. As a result, a significant portion of the historical record has become increasingly difficult to reach.

Artificial intelligence technologies have offered a partial remedy to this condition. Tasks such as script recognition, transcription, transliteration, and even semantic analysis can now be performed using digital tools. Platforms like Transleyt, Transkribus, Vesiqari, and AKİS have contributed to this transformation in different ways. Some have proven effective for printed sources, while others have specialized in manuscripts. Still, many remain in developmental stages. Determining whether their outputs meet scholarly standards ultimately depends on the historian's critical judgment.

This study has sought not only to highlight the benefits these technologies offer to historical text processing but also to underline their limitations and potential risks. AI tools may assist the historian, but they cannot replace the process of interpretation itself. In the end, they serve as mediators—not substitutes—for scholarly engagement. As these systems continue to evolve, their transparency, accessibility, and context sensitivity must improve. The future of historiography in the digital age may be enriched by artificial intelligence, but its direction remains firmly in the hands of human inquiry.

GENİŞ ÖZET

Bu makale, yapay zekâ teknolojileri ile Osmanlı araştırmaları arasındaki kesişim alanını ele alıyor ve tarihçilerin tarihî metinlerle kurduğu ilişkinin dijital çağda nasıl dönüşmekte olduğunu inceliyor. Çalışmanın merkezinde, Osmanlı Türkçesiyle yazılmış metinlerin okunması, çözümlenmesi ve yeniden inşasında dijital araçların -özellikle yapay zekâ destekli platformların- nasıl bir rol oynadığı sorusu yer alıyor. Osmanlıca belgeler, el yazmaları ve basılı kaynakların tarihçiler için uzun yıllardır temel bir araştırma malzemesi oluşturduğu su götürmez bir gerçek; ancak erken Cumhuriyet döneminde gerçekleştirilen Harf Devrimi ve bunu takip eden dili sadeleştirme politikaları, bu metinlerle modern bilim dünyası arasındaki bağı büyük ölçüde kopardı. Latin harflerine geçiş, her ne kadar okuryazarlığı artırma amacı taşısa da, dilin söz varlığını daralttı, dizilimi değiştirdi ve geçmişle kurulan kültürel sürekliliği

bizzarure zedeledi. Geoffrey Lewis'in ifadesiyle bu süreç, "trajik bir başarı"ydı. Sonuçta Osmanlı Türkçesiyle kaleme alınmış belgeler yalnızca tarihçiler için değil, beşerî ve sosyal bilimlerin farklı alanlarında çalışan araştırmacılar için de büyük ölçüde anlaşılmaz hale geldi.

Makale, sorunun sadece alfabe değişikliğinden ibaret olmadığını vurguluyor. Asıl mesele, Osmanlıca metinlerin güvenilir biçimde neşredilmesi, yani orijinal ya da orijinale en yakın nüshaların belirlenip sağlıklı tenkitli edisyonların hazırlanması olarak ortaya çıkıyor. Bu noktada Türkiye'deki tenkitli neşir geleneğinin tarihine de değiniliyor makalede. II. Meşrutiyet döneminden itibaren metin tenkidi ve tahkikli neşir alanında Avrupa'daki çağdaş örneklerle yarışan çalışmalar yapılmış; modern anlamda metin kurma ve nüsha tespiti yöntemleri uygulanmaya başlanmıştı. Makale, yaptığı atıflarla, bir metin üzerinde çalışmaya başlamadan önce yapılması gereken en önemli işin nüshaların tespiti, karşılaştırılması ve "nüsha ailelerinin" (stemma codicum) çıkarılması olduğunu hatırlatıyor. Bu anlamda, hangi nüshanın Latinize edileceği ya da dijital ortama aktarılacağı konusu, metinlerin sahihliğini doğrudan belirleyen ilmî bir tercih olarak öne çıkıyor.

Bu tarihî ve metodolojik çerçeve üzerine inşa edilen çalışma, son yıllarda Osmanlı metinlerinin dijitalleştirilmesi ve işlenmesi için geliştirilen yapay zekâ tabanlı dört platformu karşılaştırmalı olarak değerlendiriyor: Transleyt, Transkribus, Vesiqari ve AKİS. Her biri farklı bir kurum yahut teknik arka plana sahip olan bu sistemler, Osmanlı Türkçesi metinlerini Latin harflerine aktarma, karakter tanıma (OCR/HTR), anahtar kelime tarama, kişi ve yer isimlerini ayıklama gibi işlevlerde kullanılıyor. Makale, her platformun teknik altyapısını, doğruluk oranını, kullanıcı arayüzünü ve tarihî metinlerle çalışma kabiliyetini ayrı ayrı analiz ediyor.

Transleyt, özellikle matbu Osmanlıca metinlerde hızlı ve başarılı sonuçlar veriyor; Latin harfli transliterasyon ve kimi durumlarda modernize edilmiş metin çıktıları sağlıyor. Ancak bu modernizasyon, tarihî özgünlüğü korumak isteyen araştırmacılar açısından kimi zaman sorunlu bir yön taşıyor. Transkribus ise gerek matbu, gerekse matbu belgeler/eserler için geliştirilmiş bir Avrupa girişimi olarak dikkat çekiyor ve kullanıcıların kendi modellerini eğitebilmesine olanak tanıyan etkileşimli bir yapı sunuyor. Vesiqari, daha ziyade matbu (nesih) yazı tipiyle sınırlı kalıyor; özellikle rik'a el yazmalarında düşük başarı gösteriyor. AKİS ise Sabancı Üniversitesi'nin VERİM ve Dijital Beşerî Bilimler Laboratuvarı ev sahipliğinde, Medeniyet Üniversitesi ve Viyana Üniversitesi'nden araştırmacıların katılımıyla yürütülüyor. Proje, derin öğrenme tabanlı bir

sistem geliştirerek Arap harfli Osmanlı Türkçesi metinlerin (yazma ve matbu) otomatik olarak Latin harflerine çevrilmesini amaçlıyor; 1928 öncesi kaynakların beşerî ve sosyal bilimler araştırmalarında daha geniş kitlelerce erişilebilir hale gelmesini hedefleyen proje hâlen prototip aşamasında bulunan bir akademik girişim olarak öne çıkıyor. Her platform farklı düzeyde başarı gösterse de hepsinin ortak hedefi, Osmanlı metinlerinin dijital çağda yeniden erişilebilir hale gelmesini sağlamak olarak tanımlanıyor.

Makalenin önemli bir bölümü, bu dijital sistemlerin beraberinde getirdiği etik ve metodolojik meseleleri tartışıyor. En başta gelen sorun, "güven" meselesi olarak öne çıkıyor: çoğu platform, işleyişi kullanıcıya açıklanmayan "kara kutu" algoritmalarla çalışıyor. Üretilen çıktının hangi ölçütlere göre oluşturulduğu bilinmediğinde, araştırmacının ilmî sorumluluğu zedelenmiş oluyor. Ayrıca bazı platformların metinleri sadece çeviriyazı olarak vermekle kalmayıp kelime dağarcığını modernleştirdiği, hatta anlam kaymalarına neden olabileceği de öngörülüyor. Bu durumun, metinlerin özgün bağlamından kopmasına ve tarihî anlamın bozulmasına yol açabileceği gerçeği de ortada duruyor. Makale, yapay zekânın kolaylaştırıcı fonksiyonuna rağmen yorumlama yetisinin tamamen insana özgü bir beceri olduğunu vurguluyor; algoritmaların anlam, üslup ve bağlam nüanslarını kavrayamadığını belirtiyor. Bu nedenle yapay zekâ sistemleri tarihçinin yerini alan araçlar değil, onun analitik ve eleştirel düşünme kapasitesini genişleten tamamlayıcı unsurlar olarak değerlendiriliyor.

Son olarak çalışma, erişim ve eşitlik konularına değiniyor. Bazı platformların ticari, bazılarının ise yalnızca proje ortaklarına açık olması, özellikle gelişmekte olan ülkelerdeki araştırmacılar için ciddi bir "dijital eşitsizlik" oluşturuyor. Tarihî bilgi insanlığın ortak mirası olduğuna göre, bu sistemlerin açık kaynaklı ve kamu destekli olması gerektiği savunuluyor.

Sonuç olarak makale, Osmanlı Türkçesiyle yazılmış metinlerle çalışmanın hem teknik hem entelektüel bir çaba gerektirdiğini vurguluyor. Yapay zekâ tabanlı araçlar, tarihçinin işini kolaylaştırıyor; veri toplama ve çözümleme süreçlerini hızlandırıyor, fakat metinlerin anlam katmanlarını çözmek, bağlamı yorumlamak ve tarihî hakikati yeniden inşa etmek hâlâ insan zekâsı ve sezgisine ihtiyaç duyuyor. Dolayısıyla yapay zekâ, tarihçinin yerini alan değil, onun metinlerle kurduğu ilişkiyi yeniden şekillendiren ve derinleştiren bir yardımcı olarak konumlanıyor. Dijital çağın tarihyazımı, insan yorumuyla makine öğreniminin dengeli birlikteliğiyle gelişiyor.

References

Aladağ, Fatma and Elif Derin Can: "Programlamadan Yapay Zekâya Osmanlı Türkçesi için Metin Tanıma ve Otomatik Transkripsiyon," *Dijital Beşerî Bilimler ve Osmanlı Çalışmaları*, ed. Yunus Uğur, İstanbul: Vakıfbank Kültür Yayınları, 2023: 71.

Berry, David M: The Philosophy of Software: Code and Mediation in the Digital Age, London: Palgrave Macmillan, 2011.

Bodleian Library: *Early Modern Manuscripts and Paleography Resources*, Oxford: University of Oxford, 2016.

Crymble, Adam. *Technology and the Historian: Transformations in the Digital Age*. Urbana: University of Illinois Press, 2021.

Demircioğlu, Cemal: *Çeviribilimde Tarih ve Tarihyazımı*, İstanbul: Boğaziçi University Press, 2016.

Demircioğlu, Cemal: "Osmanlı Edebiyat Geleneğinde Tercüme Eserleri İncelemede Bir Yöntem Önerisi: Eyleme Odaklı Söylem Çözümlemesi", In *Âb-ı Hayâtı Aramak: Gönül Tekin'e Armağan*, 245-295. İstanbul: Yeditepe Yayınevi, 2018.

Derin, Elif: "TRANSKRIBUS ile Osmanlı Türkçesi'nin Dijitalleştirilmesi", https://www.digitalottomanstudies.com/post/kopyası-transkribus-ile-osmanlı-türkçesi-nin-dijitalleştirilmesi?lang=tr&utm_source=chatgpt.com (Erişim: 10.04.2025).

Eravcı, Bahaeddin, H. Mustafa Eravcı and Efe Can Tanrısever: "Tarih Araştırmaları ve Yapay Zeka," *AUSBD*, 2024; "Yapay Zeka ve Sosyal Bilimler Öğretim" Özel Sayısı: 10-11.

Eşer, Ahmet: "Tenkitli Metin Neşrinin Esasları", *BAİD* 14 (Aralık 2021), 562- 563.

Gregory, Ian N., and Alistair Geddes: *Toward Spatial Humanities: Historical GIS and Spatial History*, Bloomington: Indiana University Press, 2014.

Gündüz, Olgun: "Edebiyat Sosyolojisi Değişmelerinde Yaklaşım Değişimleri", In *Cumhuriyetin Yüzüncü Yılında Sosyal Bilimler: Süreklilik ve Dönüşüm Sempozyumu Bildiri Kitab*ı, edited by Bilal Güzel, Ankara: Ankara Sosyal Bilimler Üniversitesi Yayınları, 2024.

Hughes-Warrington, Marnie, and Jo Guldi: "The Historian's Role in an Age of AI", Network in Canadian History & Environment, 2024. https://niche-canada.org/2024/11/21/the-historians-role-in-an-age-of-ai/

Karaca, Taha Niyazi: *Tarihçi Geçmişi Kurgulamak: Kuram, Tasarım, İnşa*, İstanbul: Kronik Kitap, 2020.

Kitchin, Rob: *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, London: Sage, 2014.

Köksal, M. Fatih: "Metin Neşrinin Ana Esasları", *TÜBAR*, XXXI (2012 Spring), p. 182.

Lewis, Geoffrey: *Trajik Başarı: Türk Dil Reformu*, Tr. by M. Fatih Uslu. İstanbul: Çeviribilim Yayınları, 2019.

Luo, Iris: "AI Interpretation vs. Human Interpreters: What Does The Future Of Interpretation Look Like?" https://www.ecinnovations.com/blog/ai-interpretation-vs-human-interpreters-what-does-the-future-of-interpretation-look-like/ (Erişim: 15.04.2025).

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze: *Introduction to Information Retrieval*, Cambridge: Cambridge University Press, 2008.

O'Neil, Cathy: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown Publishing, 2016.

Piotrowski, Michael: *Natural Language Processing for Historical Texts*, San Rafael, CA: Morgan & Claypool, 2012.

Saraç, Ahmet Emin: "Modern Metin Tenkidi ve Tesisi Usulünün Türkiye'ye Girişi", *Türk Dili ve Edebiyatı Dergisi*, v. 63, no. 2, 2023, pp. 93-113

Stokes, Peter A: "Digital Approaches to Paleography and Book History: Some Challenges, Present and Future," *Frontiers in Digital Humanities* 2/5 (2015), 1-3.

Tasdemir, Esma F. Bilgin et al: "Automatic Transcription of Ottoman Documents Using Deep Learning", In: Sfikas, G., Retsinas, G. (eds) *Document Analysis Systems*, *DAS 2024. Lecture Notes in Computer Science*, vol 14994. Springer, Cham, 2024.

Yazar, Sadık: "Bakir Bir Araştırma Sahası Olarak Osmanlı Tercüme Geleneği", *Türk Dili ve Edebiyatı Dergisi* 60/1 (2020), 153-178.