

# Dijitalleştirme Çalışmalarında El Yazması Metinlerin Yapay Zekâ Yardımıyla Deşifre Edilmesi

## *eScriptorium* Programı Örneği

RAHİME KARAYİĞİT\*

Öz

Son yıllarda İlahiyat sahasında dijital beşerî bilimlere artan ilgi, araştırmacı ve akademisyenleri bu konuda toplantılar düzenlemeye yönlendirmiştir. Dijital İlahiyat çalışmaları üzerine düzenlenen bir dizi toplantıda araştırmacıların bu konuda bilgilendirilip disiplinler arası yeni çalışma alanlarına yönlendirilmesi hedeflenmektedir. Bu araştırma notu, el yazmalarının dijitalleştirilmesi amacıyla *eScriptorium* programının pratik kullanımını konu edinmektedir. Bu yazı, programın avantaj ve dezavantajlarını tartışmayı, aynı zamanda taranan metinlerde karşılaşılan problemlere ve bunların çözümüne dair öneriler sunmayı hedeflemektedir. Bu amaç doğrultusunda araştırmacıların bu programı ne şekilde kullanması gerektiğine dair program içerisinden örnekler ortaya konmuştur. *eScriptorium* her ne kadar makine öğretimiyle yazma eserlerin dijitalleştirilmesi amacına matuf olarak geliştirilen bir program olsa da belirli düzeylerde manuel müdahalelere ihtiyaç duymaktadır. Bu sebeple dijital beşerî bilimler sahasında mühendisler ve sosyal bilimcilerin ortak çalışma grupları oluşturularak mevcut programları geliştirmesi ve yeni programlara imza atması gerekmektedir. Programların mantığı ve yapay zekânın işleyişine vukufiyetin artması, muhtemel program yazılımlarının hata oranlarının zamanla azalmasına vesile olacaktır. Böylece süreç içerisinde doğru veri girişleriyle nete yakın veriler elde etme ihtimali kuvvet bulacaktır.

**Anahtar Kelimeler:** Dijital beşerî bilimler, yapay zekâ, dijital İlahiyat çalışmaları, el yazmaları.

\* Doktora Öğrencisi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Hadis Bilim Dalı / PhD Student, Istanbul University, Institute of Social Sciences, Department of Hadith. Istanbul, Türkiye. ORCID: 0000-0001-6175-8796 e-posta: rahime.karayigit@ogr.iu.edu.tr

DOI: 10.26570/isad.1530174 • Gelis/Received 08.08.2024 • Kabul/Accepted 28.10.2024

Atıf/Citation Karayigit, Rahime, "Dijitalleştirme Çalışmalarında El Yazması Metinlerin Yapay Zekâ Yardımıyla Deşifre Edilmesi -eScriptorium Programı Örneği-", *İslam Araştırmaları Dergisi*, 53 (2025): 189-197.

## Deciphering Manuscript Texts with the Assistance of Artificial Intelligence in Digitization Studies

### The Case of the eScriptorium Program

#### Abstract

In recent years, the growing interest in digital humanities within the field of theology has led researchers and academics to organize meetings on this subject. A series of conferences on Digital Theological Studies aims to inform researchers and guide them towards new interdisciplinary areas of study. Artificial intelligence models developed especially on Arabic texts contribute to more accurate reading of manuscripts. This article focuses on the practical use of the *eScriptorium* program for the digitization of manuscripts. Its purpose is to assess the advantages and disadvantages of the program, while also providing recommendations for solving the problems encountered in the scanned texts. In line with this objective example from within the program are presented to demonstrate how researchers should utilize it. Although *eScriptorium* is a program developed for the digitization of manuscripts through machine learning, it still requires manual intervention at certain levels. Therefore, it is necessary for engineers and social scientists to form collaborative working groups within the digital humanities field to improve existing programs and develop new ones. A deeper understanding of the logic of these programs and the functioning of artificial intelligence will contribute to the reduction of potential software errors over time. Consequently, the likelihood of obtaining near-accurate data through proper data entries will be strengthened throughout the process.

**Keywords:** Digital Humanities, Artificial Intelligence, Digital Theological Studies, manuscripts.

Dijital beşerî bilimlere yönelik İlahiyat sahasında son yıllarda artan yönelim, araştırmacıları ve akademisyenleri bu konuda bilgilendirme toplantıları düzenlemeye sevk etmiştir. 2024 yılının başlarında “Dijital İlahiyat Araştırmaları 1” eğitim programının ilki düzenlenmiş;<sup>1</sup> aynı yılın 01-15 Temmuz 2024 tarihinde “Dijital İlahiyat Araştırmaları 2” programı ile seminerlere devam edilmiştir. Bu program kapsamında yapay zekâ, OCR/HTR teknolojileri, doğal dil işleme, büyük dil modeli ve fetva mecmuaları gibi konularda dersler verilmiş, üretken yapay zekânın İlahiyat sahasındaki kullanımına dair seminerler düzenlenmiştir.<sup>2</sup> Bu derslerden biri de el yazması ve matbu eserlerin dijitalleştirilmesine yönelik *eScriptorium*<sup>3</sup> prog-

1 Program içeriği hakkında tanıtıcı bilgiler için bk. Doğan, “Uygulamalı Dijital İlahiyat Araştırmaları I Eğitim Programı Tanıtımı”, s. 257-60.

2 Dijital İlahiyat Araştırmaları 2'nin kapsamı, yapay zekânın imkânları ve sosyal bilimler/ilahiyata kazandıracığı yeni yaklaşımlar hususunda araştırmacıları bilgilendirmektir (ayrıntılı bilgi için bk. <https://www.dijitalilahiyatarastirmalari.com/about-4>).

3 *eScriptorium* programına benzer el yazması ve matbu metinlerin yapay zekâ yardımıyla dijitalleştirilmesi çalışmalarını yürüten Avrupa'da <https://www>.

ramının araştırmacılarla uygulamalı öğretimi üzerinedir. TÜBİTAK (2237-A) destekli “Dijital İlahiyat Araştırmaları 2” seminer dizisinde, OpenITI proje ekibinden Jonathan Parkers Allen tarafından *eScriptorium* eğitimi verilmiş ve katılımcıların çeşitli metinler üzerinden uygulama yapmasına imkân tanınmıştır.

Dr. Jonathan Parkes Allen, Roshan Farsça Araştırmaları Enstitüsü’nde doktora sonrası araştırmacısıdır. Osmanlı din ve kültür tarihi üzerine araştırmalar yürüten Allen, dijital beşerî bilimler alanında OpenITI (Open Islamicate Text Initiative) projesinde yer almaktadır.<sup>4</sup> OpenITI<sup>5</sup> projesi kapsamında *eScriptorium* adındaki program üzerinden Arapça, Farsça, Osmanlı Türkçesi ile yazılan metinlerin makine öğretimi teknikleri uygulanarak taranıp dijital ortama aktarılması mümkündür. Projede, el yazması ve matbu Arapça harflerle yazılı metinler, OCR (Optik Karakter Tanıma) / HTR (El Yazması Metin Tanıma) teknolojileri kullanılarak deşifre edilmektedir. Şu anda deneme sürümünde olan programın, yakın zamanda araştırmacıların açık erişimine sunulması hedeflenmektedir.

OpenITI projesinin amaçlarından biri, İslam coğrafyasında yazılmış klasik dönem metinlerinin dijitalleştirilmesidir. Arap-Fars kültüründe yazılmış metinler, çalışmanın kapsamına dahil edilmiştir. Projede erken dönem metinlerden XX. yüzyıl metinlerine kadar geniş bir veri havuzu oluşturulmuştur. Başlangıçta matbu ve el yazması metinleri kapsayan proje, daha sonra sadece el yazmalarına yönelmiştir. Öte yandan el yazması metinlerin bir kısmının içinde şerh-haşiye bulunması sebebiyle tasnif aşamasında güçlük çekildiği bir gerçektir. Burada amaçlanan şey, araştırmacının büyük bir veri setiyle karşı karşıya olduğunda tek başına deşifre etmesi güç metinleri, program yardımıyla script/karakter haline dönüştürmesini sağlamaktır.

TÜBİTAK eğitiminde bu tür eserlerin dijitalleştirilmesi için OCR ve HTR teknolojilerinin teorik arka planı hakkında bilgilendirme yapılmıştır. OCR, matbu eserleri dijitalleştirme; HTR, ise el yazması metinleri dijitalleştirme için kullanılmaktadır. El yazması metinlerde farklı renk ve yazı stilleri (eğik, düz yazı) bulunabilir. OCR’ın çalışma prensibi daha kolaydır zira bu konuda belli bir standarda sahip matbu eserlerde transkripsiyon/deşifre işlemi gerçekleştirilmektedir. El yazması metinlerde OCR teknolojisinin kullanım güçlüğü, yazarın metni yazıya aktarırken kişisel tercihlerinin ve sayfayı kullanım biçiminin çok öznel olması sebebiyledir. Buna bağlı olarak

---

transkribus.org/ sitesi vardır. Ancak bu program, Osmanlı Türkçesi ve farklı dillerde metinlerde kullanılsa da henüz Arapça dil modeline sahip değildir. Bu konudaki bilgilendirmesi için Elif Derin’e teşekkür ederim.

4 [jallen22 | School of Languages, Literatures, and Cultures \(umd.edu\)](http://jallen22@um.edu).

5 Open ITI projesi 2016 yılından itibaren bu dijitalleştirme üzerine odaklanmıştır (bk. [About \[openiti.org\]](http://about.openiti.org)).

yazının hızlı ya da yavaş yazılması, okunaklığı, kelimeler arasındaki boşlukların hepsi bu tarama sistemlerinde hata olasılığını artıran unsurlardır.

Şekil 1: El yazması eğik yazılı, kenar notları olan bir metin

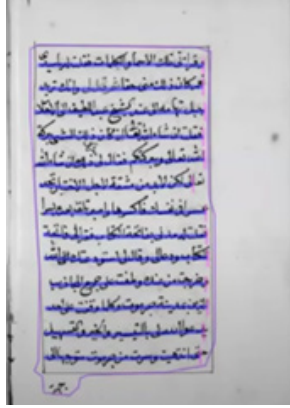


OpenITI tarafından geliştirilen *eScriptorium*, deneme sürümünden sonra araştırmacılar için erişime açıldığında el yazması metinlerin taranması sözkonusu olacaktır. Programın kullanım izni, TÜBİTAK destekli proje kapsamında verilen ders vesilesiyle dijital ilahiyat araştırmacılarına açılmıştır. Program kapsamında öncelikle XML, PDF, IIF, ZIP uzantılı belgenin<sup>6</sup> içe aktarılması (import), daha sonra sırasıyla belgenin Binarize, Segmentation ve Transcribe işlemleri yapılır. Programın matbu ve yazma eserler için ayrı ayrı olmak üzere dil paketlerinin yüklenmesi gerekmektedir. Program henüz deneme sürümünde olduğu için bu tarz işlemler dahi yavaş işlemektedir. Yeni geliştirilen sürümde dosyalar, bulut veri tabanı ile yükleneceği için verilerin yüklenmesinin hızlanması öngörülmektedir. Programa dair Youtube üzerinden yükleme ve sonrası işlemler hakkında bilgi edinilebilir.<sup>7</sup>

6 JPEG formundaki belgeler ise sürüklenerek programa entegre edilebilir.

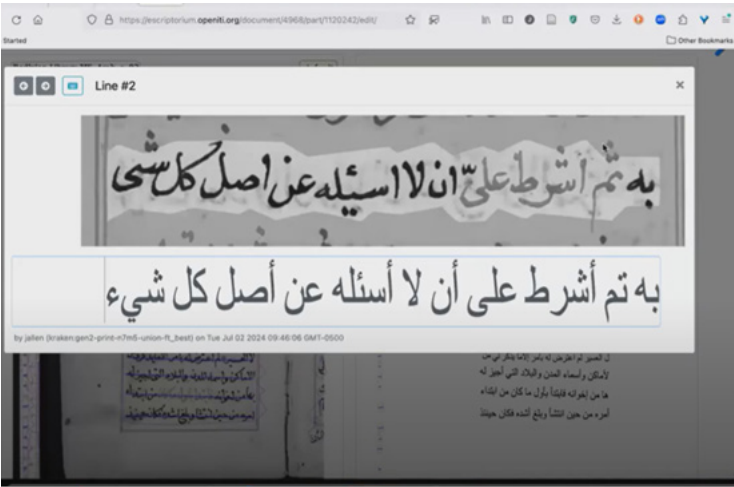
7 <https://www.youtube.com/watch?v=N0hSNC3YvD4>.

Şekil 2: Taraması düzgün yapılmış sayfanın ayrıştırılması



Şekil 2’de görüldüğü üzere program, ilk aşamada metni satırlar halinde ayrıştırarak transkripsiyon işlemi yapar. Bu tarama neticesinde aşağıda görülen odaklanılmış satırlar üzerinden metnin kontrolleri yapılır, eksiklikleri varsa giderilir.

Şekil 3: Taraması yapılmış bir metinde programın kelime okumadaki hata oranı



Şekil-3’te program tarafından okunan metinde تم ifadesinin تم şeklinde eksik tarandığı görülmektedir. Metnin sonundaki شيء ifadesinde ise aslında harfin noktaları tarama dışında kalmış olmasına rağmen program kelimeyi doğru bir şekilde okumuştur. Bu tür aksaklıklar neticesinde programda manuel düzeltme seçeneği kullanılarak ibarenin doğru bir şekilde aktarımı yapılabilmektedir.

Metnin doğru taraması bütün satırlarda doğru bir şekilde yapılmadığında, program Arap harfleriyle yazılan metni istenilen şekilde okuyamamaktadır. Şekil 4'teki resimde tarama esnasında altıncı satırda eksik bir tarama söz konusu olduğu için çizgi tamamlanmamıştır. Bu durumda kullanıcı, o çizgiyi el ile müdahale ederek uzatmalı ve metnin taranmasına imkân sağlamalıdır. Araştırmancın yürütücüleri başlangıçta bu problemi küçük ve kolay çözümler bir şey olarak düşünmüşlerse de program sürecinde bu hata kodunun düzeltilmesi oldukça vakit almıştır.

**Şekil 4: Taraması düzgün yapılmamış metnin Arapça ibarenin tamamını okuyamaması**



Program, proje kapsamında devam eden araştırma sürecinde farklı dillerdeki metinleri de taramaya başlamıştır. Aşağıda örneği verilen eğik marjine sahip metin, programın ilk kullanılmaya başlandığı dönemde taranamazken gelişmeye devam eden bu teknolojiye programa kazandırılmıştır.

**Şekil 5: Eğik marjinli ve metin dışı unsurlara sahip bir metin**

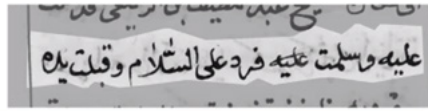


Şekil 5'teki Farsça metinde farklı yönlerle bakan eğik yazıların yanı sıra kuyudât bilgisi bulunmaktadır. Farklı şekillerde yazılmış bu metni okuyabilmesi, yapay zekânın olumlu tarafı olarak zikredilebilir. Ancak müellif yahut müstensih tarafından farklı amaçlarla sayfaya eklenmiş metinler, yapay zekâ tarafından metnin kendisi mi yoksa metin dışı unsur mu olduğu konusunda ayırt edilememektedir. Burada tekrar insan gücüne dönerek araştırmacıdan buna el ile müdahale edip değiştirmesi ve asıl metin ile metin dışı unsuru ayırt etmesi beklenmektedir.

Programın geliştirilmesi gereken taraflarından biri de metin analizi yapılırken kelimelerden bir grup, el yazısında birleşik bir şekilde yazıldıysa, metin içerisinde o kelime taratıldığında programın buna cevap verememesidir. Söz gelimi "haberunsahihun" ifadesi birleşik bir şekilde yazıldıysa program bu kelimeyi iki ayrı kelime olarak ayırtıramamaktadır. Özellikle nicel veri analizi için programı kullanmayı düşünen araştırmacıların bu durumu dikkate alması gerekmektedir.

Programda dikkat edilmesi gereken unsurlardan biri de yazma eser sahibinin yazı yazarken üzerini çizdiği, yerini değiştirdiği kelimelerin, program tarafından taranması ve bu çizginin program tarafından görülmemesidir. Şekil 6'daki yazma örneğinde kelimenin yeri başlangıçta yanlış yazılmış ardından bu hata farkedilerek üzeri çizilmiş ve doğru yere yazılmıştır. Programın bu durumu anlayamaması sebebiyle, programı kullanan araştırmacının yanlışlığa dikkat çekmek üzere manuel olarak bu kelimeyi süslü paranteze alması gerekir.

#### Şekil 6: Kelimenin sırasına yönelik dikkat



عليه وسلم {عليه} فرد على السلام وقت يد

Yapay zekânın günümüzde giderek gelişen durumunun ilmi sahadaki bir yansımaları olarak değerlendirilebilecek bu programın, henüz yapım aşamasında olması, programa dair aksaklıkların hâlâ var olduğunu göstermektedir. *eScriptorium* her ne kadar makine üzerinden yazma eserlerin dijitalleştirilmesi yönünde bir çabaya sahip olsa da programın belirli aşamalarında insan gücüne dayalı manuel müdahalelerin olması, dijital sahaya yönelik bu çabaların çokça gelişime ihtiyaç duyduğunu göstermektedir. Dijitalleştirme çalışmalarında özellikle yazma eserlerle ilgili süreçlerde editöryal tashih için hâlâ insan gücüne başvurulması gerekmektedir. Ancak yine de bu tür teknolojilerin ve programların geliştirilmesiyle iş yükünün azalması

ve yapılan işlerde daha hızlı sonuç alma, özellikle bilgisayar teknolojisinin eserler arasındaki mukayeseyi insana göre daha çabuk yapması ve analize imkân vermesi mevcut programları cazip kılmaktadır. Aynı zamanda yazma eserlerin belirli bir standartla daha az hata payına sahip bir şekilde okunabilmesi bu teknolojinin faydaları arasında zikredilmelidir. *eScriptorium*'un yazma eser üzerine çalışan araştırmacıların istifadesine en yakın zamanda sunulması, bu alanda yapılan çalışmaların kalitesine katkı sunacaktır.

### **Bibliyografya**

Doğan, Nuran Sarıcı, "Uygulamalı Dijital İlahiyat Araştırmaları I Eğitim Programı Tanıtımı", *Marmara Üniversitesi İlahiyat Fakültesi Dergisi*, 66 (2024): 257-60.

About (openiti.org) (Erişim: 13.12.2024).

<https://www.dijitalilahiyatarastirmalari.com/about-4>. (Erişim: 13.12.2024).

<https://www.transkribus.org/> (Erişim: 13.12.2024)

<https://www.youtube.com/watch?v=N0hSNC3YvD4> (Erişim: 13.12.2024).

jallen22 | School of Languages, Literatures, and Cultures (umd.edu) (Erişim: 13.12.2024).

---

---

## **Deciphering Manuscript Texts with the Assistance of Artificial Intelligence in Digitization Studies**

### **The Case of the eScriptorium Program**

#### **Extended Summary**

With recent developments in artificial intelligence interest in digitization in academic studies has increased. Thanks to artificial intelligence, processes such as data analysis and processing, automatic text transcriptions, translations, literature reviews, presentation preparations, infographics, and simulations have become significantly easier. In Islamic studies the contributions of artificial intelligence (AI) vary across different fields of knowledge. For instance, in a legal ruling related to Islamic jurisprudence (*fiqh*) AI can classify rulings from different schools of thought by scanning literature. It can compare fatwas and organize them systematically. In hadith studies AI can generate isnad (chain of transmission) diagrams and compile commentary on the hadith. It can also collate information on narrators by comparing data from relevant biographical works. In the field of *tafsir* (Qur'anic exegesis), AI can perform semantic and grammatical analyses of words in the Qur'an and provide commentaries from various exegetes regarding the context of the verses.

Another contribution of artificial intelligence to Islamic literature is the digitization and transcription of manuscript texts. The *eScriptorium* program, developed as part of the OpenITI project, uses Handwritten Text Recognition (HTR) technology to convert manuscript texts into digital script, making them accessible to researchers. This enables



the use of handwritten texts in a digital format, even if they exhibit various styles and scripts. This article presents the practical use of the *eScriptorium* program, as discussed in a seminar funded by TÜBİTAK. Its purpose is to assess the advantages and disadvantages of the program, highlight the challenges encountered during text scanning, and offer solutions to these issues. Examples from within the program are provided to demonstrate how researchers should use it effectively.

Although *eScriptorium* was developed with the goal of digitizing manuscripts through machine learning, it still requires manual intervention at certain stages. Therefore, in the field of digital humanities, engineers and social scientists need to collaborate to further develop existing programs and create new ones. Increasing familiarity with the logic behind these programs and the functioning of artificial intelligence will likely reduce error rates in software over time. As a result, more accurate data will be obtained through proper data input, leading to greater precision in research outcomes.