Implementing AI in the National Library of Spain: The ETSO Project and Stylometric Analysis of Golden Age Theatre

Alberto Peralta

Abstract The Biblioteca Nacional de España (BNE), founded in 1711, serves as Spain's national library, safeguarding one of the most significant collections of Spanish Golden Age theatre manuscripts. With a staff of 364 and an annual budget of €35.86 million, it is a cornerstone of Spain's cultural preservation efforts. In 2017, the ETSO project ("Stylometry Applied to the Theater of the Golden Age") was launched to address the lack of transcribed texts from this period. Using AI tools like Transkribus, ETSO has automated the transcription and modernisation of over 1800 theatrical works, achieving up to 97% accuracy for printed texts and 91% for manuscripts. The project is ongoing and continues to refine its methodologies while contributing to the CETSO corpus and the TEXORO search engine. The project faced challenges such as adapting AI to historical typefaces, modernising archaic spellings, and overcoming transcription errors in handwritten manuscripts. These were mitigated through extensive model training and interdisciplinary collaboration. ETSO demonstrates how AI can unlock cultural heritage, making historical texts accessible for research and public engagement and serving as a model for integrating technology into humanities projects.

Keywords Biblioteca Nacional de España (BNE) · ETSO project · Stylometry and AI transcription · Spanish Golden Age theatre manuscripts · Cultural heritage and digital humanities

A. Peralta (✉)
Department of Business and Economics, Faculty of Economics, Business and Tourism, Universidad de Alcalá, Madrid, Spain

College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates
e-mail: alberto.peralta@adu.ac.ae

I. Mergel, C. Schmidt (eds.), *AI Innovations in Public Services*,
https://doi.org/10.1007/978-3-032-01344-6_11

1 Introduction

The preservation and study of cultural heritage are central to understanding historical and societal evolution. The Biblioteca Nacional de España (BNE), established in 1711, plays a pivotal role in this endeavour by safeguarding an extensive collection of Spanish Golden Age theatrical manuscripts and printed works. These texts, dating from 1492 to 1659, offer invaluable insights into the artistic, cultural, and societal dimensions of one of Spain's most prolific literary eras. Despite their significance, a large proportion of these works remain inaccessible due to the lack of digital transcription and modernised spelling [1].

The ETSO project ("Stylometry Applied to the Theater of the Golden Age") addresses this gap by leveraging artificial intelligence to transcribe and analyse these historical texts [1–3]. Specifically, the project employs tools like Transkribus [4] and Stylo [5] to overcome challenges posed by historical typefaces, archaic orthography, and handwritten manuscripts [6–9]. The central question of this research revolves around how AI can be applied effectively to automate the transcription of complex historical texts and facilitate their analysis for authorship attribution, literary studies, and broader academic engagement.

Previous research has demonstrated the potential of AI in similar fields, such as Optical Character Recognition (OCR) for modern texts and Handwritten Text Recognition (HTR) for historical manuscripts. However, these methods have shown limited effectiveness when applied to the intricate typefaces and diverse handwriting styles of Golden Age manuscripts. The ETSO project builds upon these advancements by developing AI models tailored specifically for Spanish Golden Age texts, achieving remarkable transcription accuracy of 97% for printed texts and 91% for manuscripts.

The rationale for this project stems from the urgent need to democratise access to cultural heritage and unlock the analytical potential of previously neglected historical texts. By transcribing and modernising over 1800 works, the ETSO project enriches the CETSO corpus and the TEXORO search engine, creating a robust foundation for further philological and literary research. This endeavour aligns with the BNE's mission to promote cultural preservation and public engagement with Spain's literary legacy.

The project employs a combination of machine learning techniques and interdisciplinary collaboration. AI models were trained on a large dataset of Golden Age texts, while scholars from various fields refined the outputs through iterative processes. Key findings include the successful attribution of previously anonymous works, such as *La francesa Laura*, to renowned playwrights like Lope de Vega. These findings demonstrate the transformative potential of AI in historical research and cultural preservation.

In conclusion, the ETSO project not only addresses the technical and methodological challenges of transcribing and analysing historical texts but also

underscores the broader implications of integrating AI into the humanities. By bridging the gap between technology and cultural heritage, this initiative sets a precedent for future projects aiming to preserve and disseminate humanity's literary and historical treasures.

2 Description of the Project

Thanks to the AI-powered tool Transkribus (READ-COOP, 2024) and its Handwritten Text Recognition (HTR) techniques, Dr. Cuéllar and the ETSO team have trained three models, which are being used and refined by the research community, capable of automatically transcribing and modernising these documents with a high degree of accuracy: approximately 97% accuracy for printed texts and 91% for manuscripts. Through these models, ETSO has processed around 1800+ theatrical works contained in prints and manuscripts from numerous libraries, archives, and other digitised sources. The resulting transcriptions are now part of the CETSO Corpus and the TEXORO search engine, both part of the ETSO project. Moreover, they possess sufficient quality to undergo stylometric analysis, which reveals interesting authorship attributions.

2.1 Need Behind the Implementation of the ETSO Project

2.1.1 Problem

- An elevated percentage of printed and handwritten theatrical texts from the Spanish Golden Age period have never been transcribed into analogue or, of course, digital formats. Consequently, it is impossible to use these documents for searches of interest or for valuable computer analyses (such as stylometry, topic modelling, sentiment detection, etc.) that have been developing in recent years.
- Until ETSO, there were no automatic spelling modernisers for Spanish that could work over computer readable (i.e. transcribed) and non-readable (i.e. non-transcribed) copies of manuscripts.

2.1.2 Need

To automatically transcribe the Spanish Golden Age documents for stylometric analyses able to establish trusted authorship attributions.

2.1.3 Drivers

- A large number of documents (1200+) from the Spanish Golden Age were initially available to the IPs, collected from numerous open sources, such as AHCT, Artelope, Aula Biblioteca Mira de Amescua, BVMC, Moretianos, Teatro de Autores Portugueses do Séc. XVII, and others, as well as from other research groups, such as GRISO, ISTAE, Prolope, and others, and individuals and researchers [2].
- The discovery of the Transkribus tool [4] in the beginning of 2020 by the ETSO team (Dr. Cuéllar). The tool aims to be a solution for transcribing ancient manuscripts with a high degree of accuracy using machine learning. Developed by the READ-COOP group [7, 8], this tool features a powerful AI system for training in text recognition and, ultimately, for automatic transcriptions. It is also user-friendly and easy to learn, making it useful even for manual transcription of texts. Importantly, this tool is not limited to any specific language or alphabet; it is designed to recognise text regardless of the language or script used. The basic process involves accurately transcribing a sufficient number of words so that, later on, through an AI system with neural networks, the machine learns to link each part of the text to its corresponding transcription. By requiring tens or even hundreds of thousands of correctly transcribed words, the machine becomes capable of accurately transcribing the lines of new documents entered into it.
- The intervention of Dr. Germán Vega García-Luengos (Universidad de Valladolid) as the Coordinator of the Golden Age Theater Portal at the BNE, and the motivation of the senior management of the BNE, to incentivise the digitalisation and analysis of the Spanish Golden Age theatrical works [1, 2].

2.2 Actors Involved in the ETSO Project

The ETSO project was initiated by Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid), who played central roles in its co-design and implementation. Their teams collaborated with the research group on Lope de Vega PROLOPE [10] and over 200 researchers, students, experts, and collectors from various institutions. Patrons, including scholars and philologists, were actively involved in refining and utilising the tools, serving as evaluators and contributors to the development of models like CETSO and TEXORO. External collaborators, such as the READ-COOP developers, provided critical support for AI model training and deployment, enhancing the project's usability and effectiveness for a broad range of academic and cultural stakeholders.

Table 1 provides greater detail on who the internal actors are in the ETSO project and the activities that they developed in the project.

In addition, the ETSO project has been developed through the collaboration of a diverse and extensive group of researchers from various organisations external to

Table 1 The internal actors in the ETSO project

Value co-creation-related activity	Actors
Co-initiation	Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid) and their respective teams
Co-design	Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid), their teams and the research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona)
Co-implementation	Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid), their teams and the research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona) and 200+ researchers, students, experts, and collectors
Co-use/production moment	Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid), their teams and the research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona), and 50+ researchers from across the world
Co-evaluation	Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid), their teams and the research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona)

the Biblioteca Nacional de España (BNE). These organisations, representing a wide range of academic and cultural institutions, are listed alphabetically in Table 2. This collective effort highlights the interdisciplinary and international scope of the project, underscoring its significance in advancing the study of Spanish Golden Age theatre.

Moreover, the ETSO project has included 122 students from the Universidad de Valladolid collaborating with the team in the development and deployment of its tools.

The collaboration of the individuals from these external institutions is threefold:

1. Individuals and institutions have provided a large number of manuscripts and other documents (1200+) from the Spanish Golden Age needed to train the AI to accurately transcribe them.
2. Developers from READ-COOP group and researchers linked to this open-source platform helped train the first recognition models in Transkribus.
3. Researchers and interested individuals keep refining the CETSO and TEXORO tools [2, 3], populating the Corpus of documents and improving the search and recognition engines. In this process, researchers receive the value of the existing tools for the philological investigations conducive to PhD dissertations and evaluations of their research questions.

Table 2 The external actors in the ETSO project

Affiliation	Affiliation (continued)
Universidad Autónoma de Madrid	Università degli Studi di Milan
Universidad Complutense de Madrid	Universitá degli studi di Salerno
Universidad de Alcalá	Università degli Studi di Torino
Universidad de Burgos	Università degli Studi di Trento
Universidad de Castilla-La Mancha	Università degli studi Roma Tre
Universidad de Córdoba	Università del Piemonte Orientale
Universidad de Guadalajara	Università del Salento
Universidad de Jaén	Università di Bologna
Universidad de La Habana	Università di Firenze
Universidad de La Rioja	Università di Pavia
Universidad de León	Universitat de Barcelona
Universidad de Navarra	Universitat de Girona
Universidad de Salamanca	Universitat de les Illes Balears
Universidad de Santiago de Compostela	Universitat de València
Universidad de Sevilla	Universität Wien and Universitat Autònoma de Barcelona
Universidad de Valladolid	Université de Genève
Universidad del País Vasco	Université de Montréal
Universidad Internacional de La Rioja	Université de Neuchâtel
Universidad Nacional de Educación a Distancia	University of California
Università Ca' Foscari Venezia	University of Kentucky
Università degli studi di Cassino e del Lazio Meridionale	Uniwersytet Wrocławski
Università degli Studi di Firenze	West Virginia University

3 Challenges Faced by the ETSO Project

3.1 First Challenge: Limitations of OCR for Transcribing Ancient Spanish Manuscripts

To attempt the automatic transcription of ancient documents, there is a common process known as Optical Character Recognition (OCR). OCR is used for digitally converting modern texts with high efficiency, but it did not serve ETSO's specific purposes. The typefaces used in printing do not correspond to the typography employed today, and OCR processes, trained to recognise modern fonts, do not achieve the desired accuracy when the manuscripts are handwritten in the Spanish orthography of the fifteenth to sixteenth centuries. Moreover, in the case of ancient manuscripts, traditional OCR methods struggle to transcribe even a single word accurately (similar references can be found, for example, in [11]). Working with these automatic transcriptions would ultimately require more effort than transcribing directly from scratch, manually [12].

3.2 Second Challenge: Modernising Archaic Spellings for Corpus Compatibility

Additionally, the ETSO team needed to consider the challenging issue of spelling. To meet the ETSO's specific interests, the spelling of the works needed to be modernised, that is, updated to modern Spanish orthography. Only then can they be compared with the rest of the corpus, or body of literature already modernised by human teams, where the spelling adheres to current norms. Even if the automatic transcription process via OCR were flawless and managed to transcribe letters and words precisely, the resulting text would not serve ETSO's specific purposes because of the mismatch between current and old meanings. In other words, the transcription would be technically valid, but the words transcribed would remain in a state where they cannot be compared to other words in works currently in the corpus.

3.3 Third Challenge: Inaccessibility of Digitised Golden Age Theatre Texts for Stylometric Analysis

The vast quantity of printed and manuscript works—hundreds or even thousands—of Golden Age theatre, although digitised and housed in valuable portals such as the BNE or the Virtual Library Miguel de Cervantes, could not be used for analysis because they were not transcribed. In other words, the text could not be read or understood by the computer, making it impossible to utilise these works for stylometric analyses. Therefore, thousands of theatrical pieces that should have been included in ETSO were excluded from the project's reach.

Quoting Dr. Cuéllar while explaining the challenge ([13]: 102):

> ["In Figure 1, we can see an example of a typical page from a printed work of Golden Age theater, in this case from the play *Duelos de amor y lealtad* by playwright Calderón de la Barca. As evident, the typography, except for occasional peculiarities typical of such documents, is clear and legible for any modern reader, making transcription straightforward. However, standard computer processes are not equipped for handling this type of text."
>
> "We also observe an example in Figure 2 of typical pages from a manuscript, such as that of the anonymous work *La francesa Laura*. Immediately, the complexity of manuscripts becomes apparent—often with elevated reading difficulty or fragments impossible to transcribe even for trained historians or philologists."]

3.4 Fourth Challenge: Inaccessibility of Digitised Golden Age Theatre Texts for Stylometric Analysis

In 2020, Transkribus did not have any pre-existing models for recognising Spanish, so it was necessary to build one from scratch. The first step involved training the tool in print recognition, which led to the creation of the initial recognition model

called Spanish Golden Age Prints 1.0. This model is now available for use by the research community. Additionally, a later, refined version of the model, incorporating modernised spelling to align with current Spanish, can also be freely utilised through the READ-COOP platform (see https://readcoop.eu/model/spanish-golden-age-prints-1-0/ and https://readcoop.eu/model/spanish-golden-age-theatre-prints-spelling-modernization-1-0/).

3.5 Fifth Challenge: Developing Custom AI Models for Spanish Golden Age Texts

Even with the proper tools to recognise print from old works, different errors in the resulting texts arise due to a range of reasons: especially cumbersome handwriting, the presence of crossed-out sections, document dirtiness, the translucency of the opposite side, and the luck (or lack thereof) of having the AI tools trained with documents of similar handwriting.

3.6 Sixth Challenge: Increased Complexity and Error Rates in Manuscript Transcriptions

Manuscripts are much more challenging than printed works and present additional complexities (see [11, 14]). Consequently, there is a significant increment in the number of errors in the automatic transcriptions of this type of text.

3.7 Seventh Challenge: Ensuring Accuracy in Stylometric Analyses of Transcriptions

In the case of printed works and manuscripts, automatic transcriptions can be used for stylometric analyses. However, the results must be handled cautiously due to the error percentage obtained by the recognition tool. In ETSO's example, the automatic transcription of some manuscripts revealed a very strong stylometric relationship with the repertoire of Lope de Vega—this playwright is the focus of the project up to now. He is, together with Miguel de Cervantes, the author of El Quijote, at the culprit of Spanish literature. They used the model-generated transcription as the base text for manual corrections and repeated the tests. These confirmed the initial result and associated the work with Lope's corpus across various analyses. Additionally, an intensive philological investigation led to the attribution of at least one anonymous work to this author.

4 Results of the ETSO Project

In synthesis, the ETSO project has advanced the preservation and study of Spain's Golden Age theatre through the innovative application of AI, bridging tradition and technology. In collaboration with the National Library of Spain (BNE), the project has successfully transcribed, modernised, and analysed a vast corpus of historical texts, uncovering hidden literary treasures such as the attribution of *La francesa Laura* and *Mujeres y criados* to Lope de Vega. By leveraging tools like Transkribus [4, 8] and Stylo [5], and fostering interdisciplinary collaboration among philologists, historians, and technologists, ETSO has overcome challenges in digitisation and transcription accuracy. Its efforts have not only enhanced access to these cultural assets for scholars and the public but have also set new standards for AI-driven research in the humanities.

Through this transformative work, ETSO exemplifies how technology can enrich cultural heritage and inspire deeper engagement with history [2, 3, 13, 14]. We now present these results in greater detail.

4.1 ETSO and the National Library of Spain: Bridging Heritage and Innovation

The Golden Age theatre represents a cornerstone of Spain's cultural history and universal dramaturgy, renowned for its artistic brilliance, a prolific array of authors and works, societal reach, and enduring global influence. Its legacy continues to thrive on contemporary stages, affirming its timeless appeal.

In this context, the BNE plays a pivotal role in preserving and studying this rich heritage. As the custodian of the world's most extensive collection of manuscripts and printed materials from the Golden Age theatre, the BNE provides unparalleled insights into this era's theatrical landscape. The ETSO project complements the BNE's efforts by uncovering critical aspects of the theatre's historical context, cementing its status as an invaluable cultural heritage. Furthermore, ETSO emphasises its relevance in modern performance, academic research, and the integration of cutting-edge technologies for preservation and dissemination.

ETSO aligns with the BNE's broader strategy to digitise and make its collections accessible for research. This mission peaked with the comprehensive digitisation of the Golden Age Theater Collection, which has been systematically made available online through the Biblioteca Digital Hispánica and Hemeroteca Digital portals since 2009. As former BNE director Ms. Ana Santos Aramburu aptly stated, "We are aware that the digitisation of culture serves to create wealth and has economic value in addition to its cultural significance" [15].

In collaboration with the research group on Lope de Vega PROLOPE [10], the BNE has undertaken long-term digitisation efforts, integrated within broader initiatives such as the Recognition and Enrichment of Archival Documents (READ)

project [7, 8]. These initiatives have pioneered advancements in Handwritten Text Recognition (HTR), Key Word Spotting, Layout Analysis, and Automatic Writer Identification, setting new benchmarks for the preservation and exploration of historical texts [6, 9, 11–14]. Through these efforts, the ETSO project not only preserves the legacy of Spain's Golden Age theatre but also redefines how it is studied and appreciated in the digital age.

4.2 Unveiling Value Creation and Collaboration in the ETSO Project

The ETSO project exemplifies a transformative integration of technology and humanities, demonstrating how value creation and collaboration can uncover cultural heritage's hidden treasures. Leveraging advanced AI tools like Transkribus, the project has significantly enhanced the analysis of historical texts, enabling discoveries such as attributing La francesa Laura and Mujeres y criados to Lope de Vega. This achievement underscores the project's ability to unlock invaluable insights that traditional methods could not have accomplished alone.

Collaboration is at the heart of ETSO's success, involving diverse stakeholders, including scholars, students, collectors, and the BNE itself, under the leadership of Dr. Cuéllar. This co-creation effort has resulted in tools like CETSO, a corpus of transcribed works, and TEXORO, a search engine designed to facilitate text analysis, which together provide a robust foundation for future research.

The use of AI in the ETSO project has reduced researchers' workload by automating transcription and has also fostered a collaborative ecosystem where technology and human expertise converge. Researchers, technologists, and cultural institutions work together to refine AI models, ensuring accuracy while enabling rapid transcription of previously inaccessible texts. This co-creation process amplifies the value of AI tools, as they are continuously improved through interdisciplinary input and active engagement from a diverse network of stakeholders. By integrating technological innovation with collective human effort, ETSO exemplifies how collaboration can drive meaningful advancements in preserving and understanding cultural heritage, creating both public and academic value.

4.3 Unlocking Heritage Through AI: Lessons from the ETSO Project

The ETSO project has demonstrated that AI can facilitate large-scale projects, but its success heavily relies on initial human input. Tools like Transkribus and Stylo have significantly accelerated the transcription of historical texts, achieving high levels of accuracy—up to 97% for printed texts. However, human expertise remains

essential for training these models, correcting errors, and ensuring continuous improvement.

Collaboration across disciplines has been another cornerstone of the ETSO project. By bringing together researchers, libraries, and digital service providers, AI tools created a shared platform where philologists, historians, and technologists could contribute their expertise. This collaboration highlights the potential of AI to act as a catalyst for interdisciplinary research and innovation.

The project also revealed critical challenges in handling historical data. Optical Character Recognition (OCR), typically used for modern texts, struggled with older typefaces and handwritten manuscripts. Additionally, modernising archaic spelling to align with current linguistic norms posed significant difficulties. These challenges underscore the need for AI tools to be tailored to the specific historical and linguistic contexts they aim to address.

AI has proven to be a powerful tool for discovery in the ETSO project, uncovering previously unattributed works, such as *La francesa Laura* and *Mujeres y criados*, which were attributed to Lope de Vega. This success highlights AI's potential to unlock new insights in historical and literary research, provided it is integrated with traditional analytical methods to validate findings.

The iterative nature of AI development has been another important lesson from ETSO. The team initially developed the Spanish Golden Age Prints 1.0 recognition model, which was later refined to include modernised spelling. This process illustrates the need for ongoing adaptation and refinement in AI-driven research projects to meet evolving goals and challenges.

Lastly, the ETSO project exemplifies the co-creation of public value through AI. By transcribing and modernising vast collections of manuscripts and printed works, the project has made these historical texts accessible not only to scholars but also to the public. This aligns with the BNE's mission to preserve and democratise access to cultural heritage, demonstrating the societal and educational potential of AI in fostering broader engagement with historical assets.

5 Conclusions

The ETSO project has emerged as a ground-breaking initiative that bridges the gap between technology and the humanities, positioning the National Library of Spain (BNE) as a leader in the digital preservation and analysis of cultural heritage. By addressing the pressing challenge of transcribing and modernising Spanish Golden Age theatrical texts, ETSO has significantly advanced our ability to uncover and analyse historical treasures. This project showcases how artificial intelligence can be a powerful enabler of large-scale cultural heritage projects, fostering interdisciplinary collaboration and democratising access to historical texts (a complete summary of the ETSO project can be found in Table 3).

Table 3 BNE: ETSO project—stylometry applied to Golden Age theatre

Country	Number of employees	Type of AI solution	Year and Maturity level
Spain	667 (as of the time of writing)	Transkribus, to automatically transcribe and modernise old prints and manuscripts with a high degree of accuracy Stylo, to relate texts based on their lexical usage Other stylometric techniques based on AI to process texts	2017 Implemented and ongoing

Project description

Thanks to the AI-powered tool Transkribus (READ-COOP, 2024) and its Handwritten Text Recognition (HTR) techniques, Dr. Cuéllar and the ETSO team have trained three models, which are being used and refined by the research community, capable of automatically transcribing and modernising these documents with a high degree of accuracy: approximately 97% accuracy for printed texts and 91% for manuscripts. Through these models, ETSO has processed around 1800+ theatrical works contained in prints and manuscripts from numerous libraries, archives, and other digitised sources. The resulting transcriptions are now part of the CETSO Corpus and the TEXORO search engine, both part of the ETSO project. Moreover, they possess sufficient quality to undergo stylometric analysis, which reveals interesting authorship attributions

Need(s) behind implementation	**Actors involved**	**Challenges**
To automatically transcribe the Spanish Golden Age documents for stylometric analyses able to establish trusted authorship attributions	• Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid) • Their teams • The research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona) 200+ researchers, students, experts, and collectors from 40+ institutions and private collections	OCR limitations: Struggled with historical typefaces and handwritten manuscripts, often making manual transcription more efficient Modernising spelling: Essential to align texts with modernised corpora for meaningful comparisons Inaccessible works: Many Golden Age texts remained unavailable for analysis due to lack of transcription AI model development: Spanish Golden Age Prints 1.0 enabled accurate transcription, later refined to include modernised spelling Persistent errors: Issues like handwriting, crossed-out sections, and document quality caused inaccuracies Manuscript complexity: Handwritten texts had higher error rates than printed works Stylometry challenges: Required caution due to tool errors and potential biases in authorship attribution

Table 3 (continued)

Country	Number of employees	Type of AI solution	Year and Maturity level
Results			
Bridging Heritage and Innovation	**Unveiling value creation and collaboration**	**Unlocking Heritage Through AI**	
The Golden Age theatre, a cornerstone of Spain's cultural history, continues to thrive as a timeless artistic legacy. The BNE plays a vital role in preserving this heritage, housing the world's largest collection of Golden Age theatre manuscripts and printed works. Complementing these efforts, the ETSO project uses cutting-edge technologies to digitise, analyse, and modernise these texts, ensuring their accessibility for research and performance. In collaboration with the Lope de Vega PROLOPE group and initiatives like the READ project, ETSO has pioneered advancements in Handwritten Text Recognition and digital analysis, redefining the study and preservation of this invaluable cultural legacy	The ETSO project exemplifies the transformative potential of integrating technology and humanities, using AI tools like Transkribus to uncover cultural heritage treasures, such as attributing works to Lope de Vega. Its success is rooted in collaboration among scholars, students, and institutions, resulting in resources like the CETSO corpus and TEXORO search engine. By combining technological innovation with human expertise, ETSO reduces transcription workloads while fostering interdisciplinary co-creation, advancing the preservation and understanding of cultural heritage for public and academic benefit	The ETSO project demonstrates how AI can facilitate large-scale cultural heritage projects by accelerating transcription and analysis of historical texts, though human expertise remains essential for training models and ensuring accuracy. Collaboration across disciplines has been key, creating a shared platform for researchers and technologists to address challenges like handling historical typefaces and modernising archaic spelling. ETSO's AI tools have uncovered previously unattributed works and highlighted the importance of iterative development and refinement. By democratising access to cultural heritage, the project aligns with the BNE's mission, showcasing AI's potential to enhance both scholarly research and public engagement	

In answering the research question of how AI can be effectively leveraged to enhance the accessibility and understanding of Golden Age theatre, ETSO provides a clear and compelling response. Through innovative tools like Transkribus and the development of the CETSO corpus and TEXORO search engine, the project has transformed inaccessible manuscripts into digitised resources available for scholarly and public engagement. These efforts underscore the importance of integrating technological innovation with collaborative human expertise to address complex challenges in cultural heritage preservation.

The key lessons learned from ETSO emphasise the value of co-creation and collaboration across diverse stakeholders, including researchers, technologists, cultural institutions, and students. This synergy has not only facilitated the refinement of AI models but also demonstrated the potential for technology to unlock new avenues for literary and historical research. The project also highlights the iterative nature of AI development, where continuous improvement and stakeholder input are critical for success.

Despite its accomplishments, ETSO has faced challenges, such as adapting AI to handle historical typefaces, handwritten manuscripts, and archaic spellings. These challenges underscore the need for tailored solutions in historical research and signal opportunities for further development. Moving forward, the National Library of Spain can build on ETSO's success by expanding the scope of its digital humanities projects. Potential follow-up initiatives could include applying similar methodologies to other historical collections or enhancing public engagement through interactive digital platforms.

The outcomes of ETSO reinforce the BNE's mission to preserve and democratise access to cultural heritage. The digitisation and analysis of Golden Age theatre texts have not only preserved a critical aspect of Spain's cultural history but have also set a benchmark for integrating AI into the humanities. By creating tools and resources that serve both academic and public audiences, the project has demonstrated the transformative potential of combining technology with cultural stewardship. As the BNE looks to the future, its commitment to innovation and collaboration will be essential in continuing to unlock the value of Spain's cultural heritage for generations to come.

Acknowledgements The work for this chapter was supported by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement ID 101061516.

References

1. BNE, Bibliotecas, datos, inteligencia artificial: las nuevas rutas del conocimiento. Conference agenda (2023), https://www.bne.es/sites/default/files/repositorio-archivos/programa_jornada_ia_0.pdf and https://www.bne.es/es/agenda/bibliotecas-datos-inteligencia-artificial-nuevas-rutas-conocimiento
2. ETSO, Herramienta TRANSKRIBUS y la Estilometría aplicada al Teatro del Siglo de Oro. Presentation in Bibliotecas, datos, inteligencia artificial: las nuevas rutas del conocimiento (2023), https://www.youtube.com/watch?v=76uQN8jMenU.min12.41-40.31
3. ETSO (2024), https://etso.es/. Accessed 1 Apr 2024
4. G. Mueühlberger, L. Seaward, M. Terras, S. Ares Oliveira, V. Bosch, M. Bryan, S. Colutto, H. Déjean, M. Diem, S. Fiel, B. Gatos, T. Grüning, A. Greinoecker, G. Hackl, V. Haukkovaara, G. Heyer, L. Hirvonen, T. Hodel, M. Jokinen, P. Kahle, M. Kallio, F. Kaplan, K. Florian, R. Labahn, E.M. Lang, S. Laube, G. Leifert, G. Louloudis, R. McNicholl, J.-L. Meunier, J. Michael, E. Mühlbauer, N. Philipp, I. Pratikakis, J. Puigcerver Pérez, H. Putz, G. Retsinas, V. Romero, R. Sablatnig, J.A. Sánchez, P. Schofield, G. Sfikas, C. Sieber, N. Stamatopoulos, T. Strauß, T. Terbul, A.H. Toselli, B. Ulreich, M. Villegas, E. Vidal, J. Walcher, M. Weidemann, H. Wurster, K. Zagoris, Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. J. Doc. **75**(5), 954–976 (2019). https://doi.org/10.1108/JD-07-2018-0114
5. M. Eder, J. Rybicki, M. Kestemont, Stylometry with R: a package for computational text. R J **8**(1), 107–121 (2016)
6. L. Padró, E. Stanilovsky, FreeLing 3.0: towards wider multilinguality, in *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, (2012)
7. READ, The Recognition and Enrichment of Archival Documents (READ) project (n.d.), https://eadh.org/projects/read
8. READ-COOP (2024), https://readcoop.eu/. Accessed 1 Apr 2024
9. PRHLT (2024), HTR Website: https://www.prhlt.upv.es/handwritten-text-recognition/. Accessed 1 Apr 2024
10. UAB, *El Grupo Prolope, Colaborador de la Exposición Lope y el Teatro del Siglo de Oro* (UAB Sala de Prensa, 2018), https://www.uab.cat/web/sala-de-prensa/detalle-noticia/el-grupo-prolope-colaborador-de-la-exposicion-em-lope-y-el-teatro-del-siglo-de-oro-/em-1345667994339.html?noticiaid=1345778034284
11. P. Smith, *Handwritten Text Recognition of the Dunhuang Manuscripts: The Challenges of Machine Learning on Ancient Chinese Texts* (British Library, 2024), https://blogs.bl.uk/digital-scholarship/2024/03/handwritten-text-recognition-of-the-dunhuang-manuscripts.html
12. CoMUN-HaT, CoMUN-HaT: Contexto, multimodalidad y colaboración del usuario en procesado de texto manuscrito (2016), https://www.prhlt.upv.es/comun-hat-contexto-multimodalidad-y-colaboracion-del-usuario-en-procesado-de-texto-manuscrito/
13. A. Cuéllar, La Inteligencia Artificial al rescate del Siglo de Oro: transcripción y modernización automática de mil trescientos impresos y manuscritos teatrales. Hipogrifo (2023). https://doi.org/10.13035/H.2023.11.01.08
14. A. Cuéllar, *Cronología y Estilometría: Datación Automática de Comedias de Lope de Vega*, Texto, literatura, cultura, XXIX (Anuario Lope de Vega, 2022). https://doi.org/10.5565/rev/anuariolopedevega.483
15. A. Muñoz, Tecnología al rescate del Siglo de Oro. El Independiente (2018) https://www.elindependiente.com/tendencias/cultura/2018/11/27/tecnologia-al-rescate-siglo-de-oro/

Alberto Peralta is currently a researcher in the Department of Business and Economics, University of Alcala (Spain). He is also an assistant professor at Abu Dhabi University (UAE). Alberto does research in sustainable business model innovation, entrepreneurial economics, and value co-creation (innovation) in the public and private sectors. He also teaches strategic innovation, lean startup, and the creation of sustainable business models at several institutions.