Preserving Clarity: The MAGIC project approach to ancient manuscripts

Yahya Momtaz[1], Stefania Conte[2], Guido Russo[3]
[1] University of Naples Federico II, Italy – yahya.momtaz@unina.it
[2] University of Naples Federico II, Italy – stefania.conte@unina.it
[3] University of Naples Federico II, Italy – guido.russo@unina.it

ABSTRACT (ENGLISH)
The MAGIC project, launched in 2023, integrates advanced AI methodologies and web development to open access to the cultural assets, mainly ancient books. The project maintains access to early printed books and manuscripts through the use of high-resolution photography and bleed-through correction techniques. The project's web platform is multilingual and provides responsive design in order to make access to all the digital devices and allow global audiences to access historical documents with ease. This study explains the techniques of the MAGIC project, which connect historical and today's environments while promoting a simple and accessible digital humanities environment.
Keywords: cultural heritage; Bleed-through; React web development; digital preservation; manuscript digitization

ABSTRACT (ITALIANO)
*Preservare la chiarezza: l'approccio del progetto MAGIC ai manoscritti antichi.*Il progetto MAGIC, lanciato nel 2023, integra metodologie di intelligenza artificiale avanzate e sviluppo web per aprire l'accesso ai beni culturali. Il progetto mantiene l'accesso ai primi libri e manoscritti stampati attraverso l'uso di fotografie ad alta risoluzione e tecniche di correzione bleed-through. La piattaforma web del progetto è multilingue e offre un design reattivo per rendere accessibili tutti i dispositivi digitali e consentire al pubblico globale di accedere facilmente ai documenti storici. Questo studio spiega le tecniche del progetto MAGIC, che collegano gli ambienti storici e odierni promuovendo al contempo un ambiente di digital humanities semplice e accessibile.
Parole chiave: patrimonio culturale; Bleed-through; sviluppo web React; conservazione digitale; digitalizzazione di manoscritti

1. INTRODUCTION
Digitizing historic manuscripts is crucial for keeping hold of cultural heritage, increasing access to historical texts, and protecting them for future generations. However, various challenges remain, particularly regarding the restoration of damaged texts and maintaining their accessibility through current digital channels (Russo et al., 2020; Conte et al., 2024). The MAGIC project addresses these issues by combining advanced picture restoration techniques with a dynamic, React-based web platform to build an inclusive and interactive digital library. In order to preserve historical writings and provide access to cultural information, the project combines advanced AI approaches with accessible web development. React, a popular JavaScript library for developing interactive UI, is used to develop a web platform that supports global access to these digitized assets. Primary features include high-resolution image display via the International Image Interoperability Framework (IIIF), enabled by Mirador, a viewer designed for the exploration of digitized manuscripts. Additionally, the website offers multilingual support for diverse audiences and has a responsive design suitable for various digital devices. The research focuses on a newly started study addressing the cleanup of manuscripts affected by bleed-through, the ongoing deterioration issue caused by ink spreading through paper. This issue often harms text clarity, reducing the readability of scanned images. The MAGIC project uses a multi-phase, AI-driven approach that integrates Gaussian Mixture Models (GMMs) for segmentation and Non-Local Means (NLM) denoising for text reconstruction. This research emphasizes the combination of AI-based image processing next to recent methods to produce accessible digital copies of ancient books and manuscripts. Beyond restoring the images, this project aims to make these digital assets more accessible and usable to both the academic community and the general public through the use of appropriate technologies. In addition, it employs AI and image processing techniques to improve text readability while reducing potential harm to the source materials.

2. WEBSITE DEVELOPMENT AND FEATURES

React.js is a popular JavaScript library known for its component-based architecture. This architecture allows developers to break down the user interface into reusable, self-contained units called components, which manage their own state and logic. Components can be composed together to build complex UIs, promoting modularity, code reusability, and easier maintenance. React.js also leverages a virtual DOM (Document Object Model) to optimize performance (see Fig. 1). The virtual DOM acts as an intermediary representation of the actual DOM, allowing React.js to efficiently update only the necessary parts of the UI when changes occur (Kolomoyets & Kynash, 2023, p. 3).

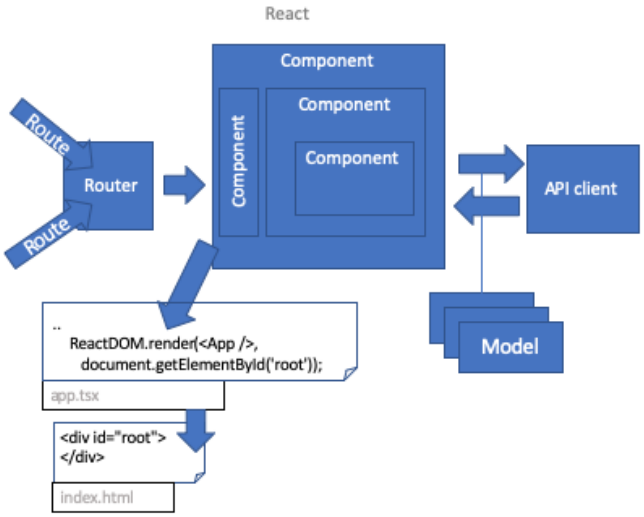

Figure 1. React architecture
(Copyright of https://handsonreact.com/docs/architecture)

The platform guarantees smooth functionality and accessibility. These architecture decisions boost the usefulness and scalability of the MAGIC platform while highlighting its dedication to preserving cultural heritage in an accessible manner for worldwide audiences. The Container and Presentational Components Pattern organizes components into logical (data-handling) and visual (UI-rendering) roles, providing clear and engaging descriptive sections for each manuscript. The Atomic Design Pattern organizes components hierarchically, facilitating fast rendering of multi-resolution JPEG 2000 images and allowing for the further analysis of text and ornaments.

Figure 2. The Magic website

The website is developed with a responsive design (see Fig. 2) and is using Bootstrap and Tailwind CSS for a consistent user experience across all devices. This responsiveness allows users to access the platform precisely from desktop computers, tablets, and smartphones. On the other hand, Framer Motion makes user interaction with the website more engaging by including smooth animations and transitions. The

platform includes Mirador, an advanced viewer for International Image Interoperability Framework (IIIF) resources, which enables users to view images at multiple resolutions, magnify small details, and look at specific sections of a text with maximum accuracy.

3. ARTIFICIAL INTELLIGENCE IN THE MAGIC PROJECT

The MAGIC project has just started using artificial intelligence (AI) to address key challenges in the protection and digitization of historical documents. A primary concern is to reduce bleed-through, a common issue when ink from one side of a page passes to the opposite side; it covers the text and decreases clarity. This issue is especially common in early manuscripts, where preservation quality deeply impacts scholarly research and public access (Conte et al., 2024). The project employs an AI-driven pipeline for bleed-through correction by improving image quality in addition to the historical integrity of the manuscripts. The procedure, now in an experimental stage, starts with segmentation, when the manuscript photos are divided into sections, separating text, background, and ornaments. Subsequently, thresholding is implemented through unsupervised machine learning methods such as Gaussian Mixture Models (GMMs) to evaluate and classify pixels according to their characteristics, which is essential for differentiating text from background noise and other components (Hanif et al., 2022). The third stage involves cleaning; it works with techniques like Non-Local Means (NLM) for denoising and inpainting to restore damaged areas by reconstructing the background texture and also to ensure a seamless restoration.

Nation	Number of Manuscripts
Vatican city	11
Italy	72
France	13
England	2
Spain	1
USA	1
Austria	1
Total	**101**

Table 1. dataset

Our dataset (see Table 1) is based on a collection of approximately 100 medieval manuscripts of Dante Alighieri's Divine Comedy that have been fully digitized over the past five years by the University of Naples. The manuscripts originate from various parts of the world, including the Vatican City, Italy, France, England, Spain, the United States, and Austria, each of which contributed one book.

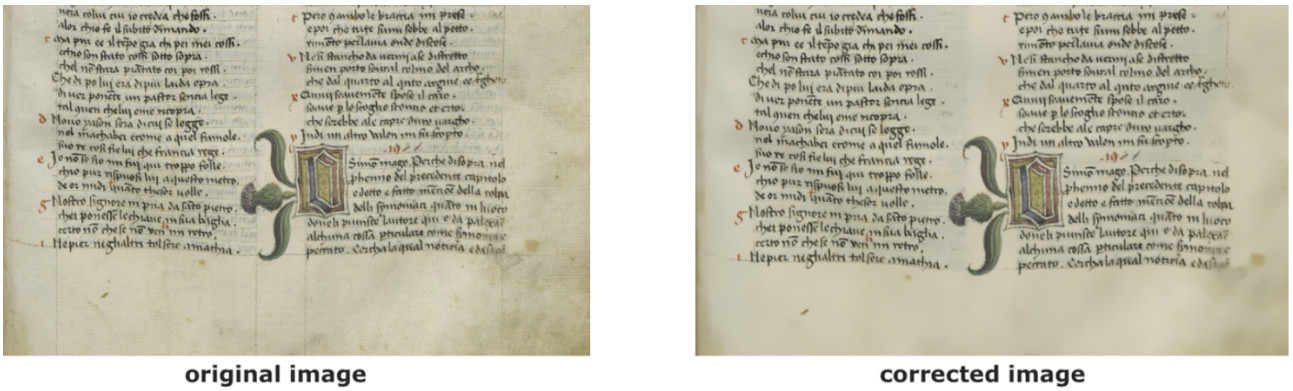

Figure 3. comparison of original image and the corrected image by bleed-through
(Original image: Firenze, Biblioteca medicea laurenziana, Plut.40.1, sheet 62r,
Corrected image is under Copyright of University of Naples Federico II, Dept. of Humanistic Studies)

These AI-driven techniques boost the visual clarity of text and their readability (see Fig. 3) for the public by combining image processing methodologies with a responsive React-based web platform. MAGIC

guarantees that restored texts are accessible to a varied, worldwide audience. The aspect of the MAGIC project highlights its value. The project integrates AI and digital humanities to tackle significant issues in manuscript preservation while promoting diversity. This method preserves cultural heritage while simultaneously expanding access to historical information for academics, educators, and the public.

4. CONCLUSION

The MAGIC project applies AI-driven image processing and React-based web development to improve historical document accessibility and preservation. The research boosts the clarity of scanned manuscripts by combining segmentation, thresholding, and cleaning methods, therefore addressing issues including bleed-through while preserving their visual and written integrity. With its simple interface, the React-based platform lets global viewers investigate these cultural gems without consider for geography.
Beyond preservation, MAGIC supports academic study, multidisciplinary cooperation, and public interaction with historical materials. This project shows how modern technology may transform digital humanities to guarantee the availability and lifetime of old knowledge for the next generations.

ACKNOWLEDGEMENTS

The project was funded by Ministry of Enterprise and Made in Italy, (code n. F/130093/03/X38 and CUP: B69J23000560005) and by Ministry of University and Research (code n. PIR01_00011, CUP: I66C18000100006).

REFERENCES

Novac, C. M., Novac, O. C., Sferle, R. M., Gordan, M. I., Bujdosó, G., & Dindelegan, C. M. (2021). Comparative study of some applications made in the Vue.js and React.js frameworks. *2021 16th International Conference on Engineering of Modern Electric Systems (EMES)*, Oradea, Romania, pp. 1–4. https://doi.org/10.1109/EMES52337.2021.9484149

Kolomoyets, M., & Kynash, Y. (2023). Front-End web development project architecture design. *Proceedings of the Conference on Computer Science and Information Technology (CSIT)*, pp. 1–5. https://doi.org/10.1109/CSIT61576.2023.10324238

Hanif, M., Tonazzini, A., Hussain, S. F., Habib, U., Salerno, E., Savino, P., & Halim, Z. (2022). Blind bleed-through removal in color ancient manuscripts. *Multimedia Tools and Applications, 82*(8), 12321–12335. https://doi.org/10.1007/s11042-022-13755-6

Russo, G., Aiosa, L., Alfano, G., Chianese, A., Cornevilli, F., Di Domenico, G. M., Maddalena, P. M., Mazzucchi, A., Muraglia, C., Russillo, F., Salvi, A., Spisso, B., Trombetti, G., & Zollo, G. (2020). MA.G.I.C.: Manuscripts of Girolamini in cloud. IOP Conference Series: Materials Science and Engineering, 949, 012081, 1–8. https://doi.org/10.1088/1757-899X/949/1/012081

Conte, S., Mazzucchi, A., Russo, G., Tortora, A., & Tortora, G. (2024). The organization and management of the MAGIC project for ancient manuscripts digitization: connections between Mediterranean cultures. *MeTe Digital: Mediterranean Networks Between Texts and Contexts. Proceedings of the XIII AIUCD Conference*, Catania, Italy, pp. 38–42.

Conte, S., Ferrante, G., Laccetti, L., Mazzucchi, A., Momtaz, Y., & Tortora, A. (Eds.). (2024). Content representation and analysis: The Magic Project and the Illuminated Dante Project integrated systems for multimedia information retrieval. In K. Roitero, M. Viviani, E. Maddalena, & S. Mizzaro (Eds.), Proceedings of the 14th Italian Information Retrieval Workshop (pp. 66–69). CEUR Workshop Proceedings.