# (idw)

### **Press release**

### Technische Universität Berlin

#### Stefanie Terp

10/24/2024 http://idw-online.de/en/news841838

Research projects History / archaeology, Information technology, Social studies transregional, national

## **Explainable AI Illuminates the Course of History**

#### Analysis shows an enrichment of old concepts through innovations

Understanding the evolution and dissemination of human knowledge over time is a long-cherished dream of many historians. A dream that has faced many challenges due to the abundance of historical materials and limited specialist resources. However, the digitization of many historical archives presents new opportunities for AI-supported analysis. Researchers from the Berlin Institute for the Foundations of Learning and Data (BIFOLD) and the Max Planck Institute for the History of Science used machine learning and explainable AI techniques to advance the historical analysis of the "Sacrobosco Collection." Their findings have now been published in Sciences Advances.

The "Sacrobosco Collection" consists of 359 early modern printed editions of astronomy textbooks from European universities (1472–1650), totaling 76,000 pages. "We developed an unsupervised machine learning model that assists the analysis of historical sources beyond human capacities through our atomization-recomposition approach," explains Matteo Valleriani, professor at the Max Planck Institute for the History of Science and BIFOLD Fellow. "Our analysis uncovers temporal and geographic patterns in knowledge transformation. We highlight the significant role of astronomy textbooks in shaping a unified mathematical culture, driven by competition among educational institutions and market dynamics."

Since antiquity, and especially during the late Middle Ages and the early modern period, the mathematical aspects of astronomy were represented in the form of numerical tables. A computational astronomical table can be understood as an expression of a modern mathematical formula, with columns displaying input values and corresponding output values. Given the significance of astronomy in the education, culture, and daily life of these epochs, the quantity of tables available for historical investigation is vast. However, the high heterogeneity of how the "same table" could be conceived, calculated, and presented complicates the investigation of these fundamental resources, rendering it, at scale, often practically impossible.

"Analysis of historical data at large presents very unique challenges from a machine learning perspective because of the extensive heterogeneity and sparseness regarding data and labels," explains Professor Klaus-Robert Müller, BIFOLD co-director and head of the Machine Learning Group at TU Berlin. "We developed the atomization-recomposition method that leverages compositional structure to achieve learning in low-resource settings, enabling an unsupervised machine learning analysis supported by explainable AI techniques."

In their approach, the researchers used an initial atomization to break down the composition of numerical features into their basic components, e.g. the task of detecting the number '15' is decomposed into detecting digits '1' and '5' respectively. From a machine learning perspective, this approach helps to efficiently model the high variety in layouts, fonts, and styles, while requiring fewer labeled annotations. A subsequent recomposition step provides the possibility to include expert knowledge and design relevant features necessary to solve the final task. For the table pages in the Sacrobosco Collection, this resulted in interpretable bigram feature maps that highlight the presence of specific bigrams, such as '15,' which aids in representing more complex numbers like '1547.' Detecting often hundreds of these



# (idw)

bigram features results in a numerical fingerprint for each page, enabling the retrieval of semantically similar content from other publications. "Our machine learning-based approach deepens our understanding by grounding insights in historical context, integrating with traditional methodologies like close reading," explains BIFOLD researcher and first author Dr. Oliver Eberle.

Following this procedure, two specific case studies were developed. The first examines the division of what was considered the habitable zones of the planetary surface into climate zones. The second case study focuses on what the researchers refer to as the Sun-Zodiac tables, which display the values necessary to determine the position of the Sun over the Zodiac throughout the year. "Overall, the historical results show that there has not been a scientific revolution, but rather a validation and an innovative enrichment of the old conceptions. This is a particularly relevant result for the history of science as a whole," concludes Matteo Valleriani.

More information concerning the research results can be found here: https://www.bifold.berlin/news-events/news/view/news-detail/explainable-ai-illuminates-the-course-of-history.

Publications: https://www.science.org/doi/10.1126/sciadv.adj1719 https://sphaera.mpiwg-berlin.mpg.de/publications/

Project: https://sphaera.mpiwg-berlin.mpg.de

Database: http://db.sphaera.mpiwg-berlin.mpg.de/resource/Start

Further information: Prof. Dr. Klaus-Robert Müller TU Berlin/BIFOLD Machine Learning Group Email: klaus-robert.mueller@tu-berlin.de

Prof. Dr. Matteo Valleriani Max Planck Institute for the History of Science Email: valleriani@mpiwg-berlin.mpg.de