

# AI to decipher ancient Indian manuscripts

Society has 52,000 rare manuscripts, many of which still have not been deciphered

Sanjay Kumar New Delhi

**The Asiatic Society in Kolkata is using AI transcription and machine learning to decipher ancient manuscripts in its archives and make them accessible to scholars worldwide.**

Founded in 1784, during British colonial rule, the Asiatic Society is one of India's oldest research institutions and is dedicated to the study and preservation of history, culture, and languages.

Many of the society's more than 52,000 rare manuscripts and historical documents have not previously been deciphered. The society launched its Vidhvanika ("decoding knowledge") project in December to digitize them and to develop language models for ancient scripts.

"Work needs to be done on the majority of the manuscripts," Anant Sinha, administrator of the Asiatic Society, Kolkata, told Arab News. "We are working with three scientists. Besides that, I have my reprography team involved in the scanning, and then there's the expert team, which includes specialists in different languages, scripts, and subjects."

The project is also being supported by the Center for Development of Advanced Computing, India's premier IT research and development organization.

The society's manuscript collection spans a wide range of subjects — including Indian history, literature, philosophy, religion, astronomy,



Students and young scholars visit the Asiatic Society's manuscript collection in Kolkata, April 4, 2024. Supplied

mathematics, medicine, and art — and of languages, including Sanskrit, Arabic, Persian, Tamil, Bengali, and other regional languages of India.

Decoding the manuscripts requires an understanding of the scripts, their language, the styles used in historical documents, the historical context, and the subject matter. There are few active, specialized paleographers and manuscript scholars conducting such work and research, not only in India but across the world.

"The motive behind this project is very simple and clear: The language, the script and the subject — generally you require knowledge of these three to understand a manuscript, (and) the people who have (that) knowledge are very few. We are developing machine language (models), so that you can use software or an app to read the manuscripts," Sinha said.

He estimated the current accuracy of the models at about 40 percent, as the machine learning process continues.

"Our plan is to take it to 90 percent to 95 percent. It will never have 100 percent accuracy," Sinha said. "It is a

## HIGHLIGHT

**Decoding the manuscripts requires an understanding of the scripts, their language, the styles used in historical documents, the historical context, and the subject matter.**

machine, it's not a human. It's learning what you are teaching it, so you have to give that leeway ... It will be an ongoing process because the machine language (model) keeps improving itself."

The Vidhvanika project was launched on the 225th

anniversary of the birth of James Prinsep, an English scholar and a former secretary of the society who is credited with deciphering the Kharosthi and Brahmi scripts of ancient India.

That feat played a crucial role in uncovering the history of the ancient Mauryan Empire that ruled over much of the Indian subcontinent during the 4th century BCE.

Vidhvanika, Sinha believes, may help save other languages that played a role in the region's history from being forgotten.

"We must make an effort to understand what is in those manuscripts and what our ancestors have left for us," he said. "Brahmi and Kharosthi are languages of this continent, and we ourselves have forgotten that. If we (are again at risk of losing) some script or some language, then we will require another James Prinsep to decipher it."