

OCR AND BEYOND

Doing Philology, Building Structured Corpora, and Computer-aided Annotation Workflows in the Age of Generative AI

28-29.03.

JST Time zone

LOCATION: Tohoku University
Katahira Campus WPI- AIMR Main Building (B01)
Seminar Room on the 2nd floor

SATURDAY

10:00-10:05
10:05-10:35
10:35-11:05
11:05-11:35
11:35-12:20
12:20-13:20
13:20-13:50
13:50-14:20
14:20-14:50
14:50-15:35
15:35

OPENING: WELCOME
Sebastian Nehrdich
On Reading Śāstra with Claude: Practical Reflections on AI in Sanskrit Studies
Kei Kataoka
Indological Research in 2026: How AI is helping, with various examples
Kengo Harimoto
“They build a hut for the sacrificer” – Towards a word-aligned corpus of Vedic prose
Oliver Hellwig
Discussion on Session 1 themes
Lead: Kyoko Amano
LUNCH
The Vedic Prose Corpus: Progress, Challenges, Visions
Sven Sellmer
eText collections, repositories, and distribution: Observations towards quality control
Julian Schott
Towards a Sustainable Dataset of Colophons, Paratexts, and Inscriptions
Ryan Conlon
Discussion on Session 2 themes
Lead: Kengo Harimoto
CLOSING

Session 1

Session 2

SUNDAY

10:30-11:00
11:00-11:30
11:30-12:00
12:00-13:00
13:00-13:30
13:30-14:00
14:00-14:15
14:15-15:00
15:00

HTR for Nepalese Manuscripts in Pracalit Script
Alexander J. O’Neill
Who Is Speaking? Encoding and Detecting Quotation Layers in Japanese Esoteric Buddhist Texts
Gaetan Rappo
Discussion on Session 3 themes
Lead: Oliver Hellwig, Sven Sellmer
LUNCH
A Preliminary Investigation of OCR Techniques for Tibetan Materials Preserved at Tohoku University
Ryuta Kikuya
Seeing the Full Elephant: Vision Language Model and LLM Use at the Dharmamitra Project
Sebastian Nehrdich
The Significance of Buddhist Studies in East Asian/Japanese Text Encoding
Kiyonori Nagasaki
Discussion on Session 4 themes
Lead: Oliver Hellwig, Sven Sellmer
CLOSING

Session 3

Session 4

REGISTER HERE
FOR ZOOM WEBINAR



CONTACT: nehrdich@tohoku.ac.jp