

Sebastian Nehrdich

Center of Integrated Japanese Studies
Tohoku University
Aoba-ku, Kawauchi 27-1
980-8576 Sendai, Japan
Phone: +81-22-795-3822
Email: nehrdich@tohoku.ac.jp

Current position

Distinguished Assistant Professor, Center for Integrated Japanese Studies, Tohoku University

Areas of specialisation

Natural Language Processing, Large Language Models (LLMs), deep learning and database construction with focus on traditional Buddhist languages such as Pāli, Sanskrit, Chinese, Tibetan, and Japanese as well as modern research languages such as English, Japanese, French, and German.

Research Experience

Oct 2025-present	Distinguished Assistant Professor, Center for Integrated Japanese Studies, Tohoku University
Apr 2023-Sept 2025	Associate Specialist, CTO of the Dharmamitra project, under the guidance of Prof. Kurt Keutzer, Berkeley AI Research Lab, University of California, Berkeley
Feb 2021-Mar 2023	Research Assistant, Heinrich-Heine-Universität Düsseldorf. Development and implementation of deep learning models and managing the technical infrastructure for the project “ChronBMM: Dating text corpora using Bayesian Mixture Models” led by Dr. Oliver Hellwig.
Aug 2019-Mar 2023	Contract work at the Khyentse Center for Tibetan Buddhist Textual Scholarship, Universität Hamburg: Design and implementation of the BuddhaNexus database for research on parallel passages in Ancient Buddhist corpora.
May-Nov 2020	Research Assistant at the Khyentse Center for Tibetan Buddhist Textual Scholarship, Universität Hamburg for the BuddhaNexus project.

Education

- Mar 2023-
Sept 2025 PHD (*summa cum laude*) at the Institute for Language and Information, Heinrich-Heine Universität Düsseldorf. Title: "Development of a Framework for the Intertextual Analysis of Indian Abhidharma and Yogācāra Literature".
- Oct 2017-
May 2020 MASTER OF ARTS in Buddhist Studies, Universität Hamburg. Title: "The Fourth Chapter of the *Madhyāntavibhāṅgikā*: Critical Edition of the Sanskrit/Tibetan and Partial English Translation". *Passed with distinction*
- Oct 2018-
Mar 2020 Exchange student at the University of Kyōto, Japan. All classes held in Japanese.
- Oct 2012-
Sep 2017 BACHELOR OF ARTS in Indology and Sinology, Universität Hamburg. Title: "Two Commentaries on Vasubandhu's *Triṃśika*: A comparative analysis of Sthiramati's and Xuanzang's exegesis".
- Sep 2016-
Jun 2017 Exchange student at the Dharma Drum Institute for Liberal Arts, Republic of China, Taiwan. All classes held in Chinese.

Grants, honours & awards

- Aug 2026 2026 Aming Tu Prize (杜正民教授獎) for Digital Buddhist Studies, awarded to the Dharmamitra project by the Library and Information Center, Dharma Drum Institute of Liberal Arts, Taiwan. The triennial prize recognizes outstanding contributions to digital research infrastructure for Buddhist Studies; previous laureates are the Buddhist Digital Resource Center (2023) and Gandhari.org (2020).
- 2026 Principal Investigator, AI for Science Exploratory Research Program (SPReAD, 萌芽的挑戰研究創出事業), Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, first call: "A High-Precision Buddhist Chinese Parallel Corpus and AI Analysis Infrastructure from the Kokuyaku Issaikyō".
- 2025 Dissertation selected for the Best Dissertation of the Year 2025 award (Beste Dissertation des Jahres 2025) by the Faculty of Arts and Humanities at the University of Düsseldorf.
- Dec 2025 Co-Principal Investigator, 2025 Humanities and AI Virtual Institute (HAVI) award from Schmidt Sciences for the project "Connectivity and Individuality in Textual Traditions: Augmenting Retrieval for Eurasian Languages".
- 2024 Google Cloud Research Credits Grant for use of Vertex AI for Cultural Preservation Tasks at the Dharmamitra project.
- 2020 Khyentse Foundation Award for Excellence in Buddhist Studies awarded by the Khyentse Foundation and the Department of Indian and Tibetan Studies, Universität Hamburg, for recognition and distinction in the field of Indo-Tibetan Buddhist Studies for the year 2020.
- 2015-2020 Scholarship for bachelor and master degree awarded by the German Academic Scholarship Foun-

ation (Studienstiftung des deutschen Volkes).

Board membership & academic service

Nov 2025-present Board of Directors, Buddhist Digital Resource Center (BDRC).

ORGANIZED EVENTS AND CHAIRED SESSIONS

- Sep 2026 Session Chair, Session 5B “Research Infrastructure and Data Curation”, JADH 2026, 15th Annual Conference of the Japanese Association for Digital Humanities, Kyushu University, Fukuoka, Japan.
- Aug 2026 Panel Co-chair (with Marieke Meelen, University of Cambridge), panel “Advances in Tibetan NLP”, 17th Seminar of the International Association for Tibetan Studies (IATS 2026), online.
- Jul 2026 Organizer and lead speaker, online workshop “Translation in the Age of AI: Dharmamitra for Translator Workflows”, a collaboration between the Dharmamitra project and the Tsadra Foundation.
- Mar 2026 Organizer, international workshop “OCR and Beyond: Philology and Digital Archives in the Age of AI”, Center for Integrated Japanese Studies (CIJS), Tohoku University, Sendai, Japan (March 28–29, 2026; co-hosted with Prof. Ryuta Kikuya).
- Dec 2025 Co-organizer, workshop and symposium “Building the Foundations of Buddhist Philology through Digital Humanities: Exploring the Potential of the Tohoku University Digital Archives (ToUDA)”, CIJS, Tohoku University, Sendai, Japan.
- Oct 2024 Co-organizer, two-day workshop “AI and the Future of Buddhist Studies”, Numata Center for Buddhist Studies, University of California, Berkeley, USA (October 18–19, 2024). Sponsored by the Glorisun Global Buddhist Network and the Ting Tsung and Wei Fong Chao Foundation.
- Jan 2023 Convenor (with Carsten Krause), hybrid workshop and international symposium “Perspectives of Digital Humanities in the Field of Buddhist Studies”, Numata Center for Buddhist Studies and Khyentse Center for Tibetan Buddhist Textual Scholarship, Universität Hamburg, Hamburg, Germany (January 12–14, 2023). Twenty speakers; workshop units on TACL, social network analysis, and BuddhaNexus.

Publications & Talks

PUBLICATIONS

- 2026 Sebastian Nehrlich, David Allport, Jivnesh Sandhan, Manoj Balaji Jagadeeshan, Sujeet Kumar, Sven Sellmer, Pawan Goyal, Kurt Keutzer (2026): Mitrasamgraha: A Comprehensive Classical Sanskrit Machine Translation Dataset. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*, colocated with EMNLP 2026. Budapest, Hungary.
- 2026 Sebastian Nehrlich, Kurt Keutzer (2026): MITRA-MT: Continued Pretraining, Parallel Corpus Mining, and Multi-Directional Machine Translation for Sanskrit, Tibetan, and Buddhist Chinese. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*, colocated with EMNLP 2026. Budapest, Hungary.
- 2026 Oliver Hellwig, Sven Sellmer, Sebastian Nehrlich, Paul Widmer, Rico Sennrich (2026): Sanskrit Word Sense Disambiguation Based on Lexicographic Definitions. In *Proceedings of the 8th International Sanskrit Computational Linguistics Symposium (ISCLS)*. IIT Roorkee, India.
- 2025 Sebastian Nehrlich, Avery Chen, Marcus Bingenheimer, Lu Huang, Rouying Tang, Xiang Wei, Leijie Zhu, Kurt Keutzer (2025): MITRA-zh-eval: Using a Buddhist Chinese Language Evaluation Dataset to Assess Machine Translation and Evaluation Metrics. In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*. Albuquerque, USA. 129–137.
- 2024 Sebastian Nehrlich, Oliver Hellwig, Kurt Keutzer (2024): One Model is All You Need: ByT5-Sanskrit, a Unified Model for Sanskrit NLP Tasks. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (Findings)*.
- 2024 Marieke Meelen, Sebastian Nehrlich, Kurt Keutzer (2024): Breakthroughs in Tibetan NLP & Digital Humanities. *Revue d'Etudes Tibétaines*, 72, 5-25.
- 2023 Oliver Hellwig, Sebastian Nehrlich, Sven Sellmer (2023): Data-driven dependency parsing of Vedic Sanskrit. *Language Resources and Evaluation*, 57, 1173-1206.
- 2023 Sebastian Nehrlich, Marcus Bingenheimer, Justin Brody, Kurt Keutzer (2023): MITRA-zh: An efficient, open machine translation solution for Buddhist Chinese. In *NLP4DH*.
- 2023 Sebastian Nehrlich (2023): Observations on the Intertextuality of Selected Abhidharma Texts Preserved in Chinese Translation. *Religions*, 14(7), 911.
- 2022 Sebastian Nehrlich, Oliver Hellwig (2022): Accurate Dependency Parsing and Tagging of Latin. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, LT4HALA 2022*. Marseille, France. 20-25
- 2022 Sebastian Nehrlich (2022): SansTib, a Sanskrit - Tibetan Parallel Corpus and Bilingual Sentence Embedding Model. In *Proceedings of the 13th Conference on Language Resources and Evaluation, LREC 2022*. Marseille, France. 6728-6734

- 2021 Oliver Hellwig, Sven Sellmer, Sebastian Nehrdich (2021): Obtaining More Expressive Corpus Distributions for Standardized Ancient Languages. In *Proceedings of the Conference on Computational Humanities Research, CHR2021*. Amsterdam, The Netherlands. 92–107
- 2020 Sebastian Nehrdich (2020), A Method for the Calculation of Parallel Passages for Buddhist Chinese Sources Based on Million-scale Nearest Neighbor Search. In *Journal of the Japanese Association of the Digital Humanities - Marcus Bingenheimer, Christian Wittern, Jinhua Chen (Guest Editors) Special Issue: Buddhism and Technology*. Vol. 5. ISSN: 2188-7276.
- 2018 Oliver Hellwig, Sebastian Nehrdich (2018): Sanskrit Word Segmentation Using Character-level Recurrent and Convolutional Neural Networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium. 2754–2763.

INVITED TALKS, WORKSHOPS, AND SEMINARS

- Sep 2026 Poster (with Tianyuan Huang, Tohoku University), JADH 2026, 15th Annual Conference of the Japanese Association for Digital Humanities, Kyushu University, Fukuoka, Japan: "Tying the Knot Between Archive- and Question-Driven Research: A Union between Two Models for Sustainable AI Integration in Japanese Studies"
- Sep 2026 Workshop, ECAI Workshop 04 (Digital Humanities), PNC 2026 Annual Conference and Joint Meetings, Dharma Drum Institute of Liberal Arts, New Taipei City, Taiwan: "Dharmamitra Workshop"
- Sep 2026 Talk, PNC 2026 Annual Conference and Joint Meetings, Dharma Drum Institute of Liberal Arts, New Taipei City, Taiwan: "Finding Friends Across Multiple Languages: Dharmamitra as a Multilingual Discovery Tool for Buddhist Textual Research"
- Aug 2026 Talk, Panel "Advances in Tibetan NLP", 17th Seminar of the International Association for Tibetan Studies (IATS 2026), online: "Multilingual Information Retrieval for Semantic Search, Retrieval-Augmented Generation, and Machine Translation of Tibetan"
- Aug 2026 Invited Plenary Presentation, Plenary Session "The Future of Tibetan Translation in the Age of AI", 17th Seminar of the International Association for Tibetan Studies (IATS 2026), online: "Dharmamitra: AI Infrastructure for Buddhist Texts"
- Aug 2026 Lightning Talk and Poster (SPReAD 1000), DiHuCo Hub Symposium, ROIS-DS Center for Open Data in the Humanities (CODH), Hitotsubashi Hall, Tokyo, Japan: "A Buddhist Chinese Parallel Corpus and AI Infrastructure from the Kokuyaku Issaikyō"
- Jul 2026 Talk, Mini-conference "From Palm Leaves to Neural Networks" (AI4AS), DH2026: Engagement, Annual Conference of the Alliance of Digital Humanities Organizations, Daejeon, South Korea: "Multilingual Semantic Search and Retrieval-Augmented Generation for Philology"

- Jul 2026 Online Workshop (with the Tsadra Foundation): "Translation in the Age of AI: Dharmamitra for Translator Workflows"
- Jun 2026 Invited Presentation, Meeting "The Possibilities of Using AI in the Translation of Buddhist Scriptures", Bukkyō Dendō Kyōkai (BDK), Tokyo, Japan: "The State of AI in Translation for Buddhist Studies: What Dharmamitra Can Contribute"
- May 2026 Invited Special Lecture, 67th Annual Conference of the Association for Indology and Study of Religion (印度学宗教学会), Tohoku University, Sendai, Japan: "Digital Tools for Indology and Buddhist Studies in the Age of AI: Recent Developments and Future Perspectives"
- May 2026 Invited Talk, Institute for Indology and Tibetology, Ludwig-Maximilians-Universität München, Munich, Germany: "Dharmamitra in 2026: Current Capabilities and Future Developments"
- May 2026 Workshop, Leipzig, Germany: "Artificial Intelligence in Buddhist Studies"
- May 2026 Invited Talk, Ahimsā Lecture Series, Asia-Africa Institute, Universität Hamburg, Hamburg, Germany: "Access as Nonviolence: Designing DH Platforms That Reduce Inequality"
- Apr 2026 Invited Workshop (full day, held online), Kathmandu University Centre for Buddhist Studies at Rangjung Yeshe Institute, Kathmandu, Nepal: "The Various MITRA Tools for Translation and Philological Research in Buddhist Studies"
- Mar 2026 Invited Talk, Faculty of Arts and Philosophy, Ghent University, Ghent, Belgium (CIJS/Tohoku University workshop "Japanese Studies in Japan and Belgium", co-organized with the Ghent University Institute for Japanese Studies): "Dharmamitra: A Platform to Support Research across Language Boundaries on Buddhist Textual Material"
- Mar 2026 Invited Guest Lecture, Ghent Centre for Buddhist Studies, Ghent University, Ghent, Belgium (Gandhāra Corpora Lecture Series): "DharmaNexus as a Multilingual Graph of Buddhist Intertextuality: Design Choices, Research Uses, and Future Applications"
- Mar 2026 Invited Talk, International Institute for Asian Studies, Leiden University, Leiden, Netherlands: "Is This the End of (Buddhist) Philology as We Know It? If So, What's Next?"
- Mar 2026 Talk, 8th International Sanskrit Computational Linguistics Symposium (ISCLS), IIT Roorkee, India: "Sanskrit Word Sense Disambiguation Based on Lexicographic Definitions"
- Jan 2026 Talk, 30th Information and Knowledge Science Forum, Doshisha University Osaka Satellite Campus, Osaka, Japan: "Dharmamitra: A Data-Driven Platform for the Research of Buddhist Texts in Multiple Languages Using Advanced NLP Methods"
- Jan 2026 Talk, 32nd Annual Conference of the Association for the Study of the History of Indian Thought, Kyoto University, Kyoto, Japan (co-presented with Kengo Harimoto, University of Naples "L'Orientale"): "AI and Indological/Buddhological Researches: Dharmamitra/DharmaNexus and Its Application"

- Dec 2025 Invited Talk, Buddhist Studies and Digital Humanities International Symposium (仏教学とデジタル・ヒューマニティーズ国際シンポジウム), Tokyo International Forum, Tokyo, Japan: "Translation, OCR, and Semantic Retrieval: Current Status and Future Outlook of the Dharmamitra Ecosystem"
- Dec 2025 Panel Presentation, 11th Symposium of Humanistic Buddhism, Taiwan (Panel "AI in the Fo Guang Dictionary of Buddhism English Translation Project and MITRA"): "Dharmamitra: A Platform that Makes Translation and Discovery of Buddhist Texts Possible Across Language Barriers"
- Dec 2025 Workshop and Symposium, Center for Integrated Japanese Studies (CIJS), Tohoku University, Sendai, Japan: "Building the Foundations of Buddhist Philology through Digital Humanities: Exploring the Potential of the Tohoku University Digital Archives (ToUDA)"
- Nov 2025 Invited Talk, Seoul National University AI Digital Humanities Center Overseas Researcher Invitation Forum, Seoul, South Korea: "From OCR via Machine Translation to Semantic Search: The Dharmamitra AI Stack for Multilingual Buddhist Philology"
- Oct 2025 Goodman Lecture Series, Khyentse Foundation, Online: "Buddhist Philology and AI"
- Sep 2025 Invited Talk, National Taiwan University, Taipei, Taiwan: "數位人文與佛教語言資源：閱讀經典的新可能 (Digital Humanities and Buddhist Language Resources: New Possibilities for Reading the Classics)"
- Sep 2025 Workshop Presentation, "解碼佛典智慧：從 AI 語言模型到數位人文分析" Workshop, Institute of Chinese Literature and Philosophy, Academia Sinica, Taipei, Taiwan: "Dharmamitra & DharmaNexus: A New Set of Digital Tools for the Study of Buddhist Texts"
- Aug 2025 Invited Talk, ELTE BTK, Kodály terem, Budapest, Hungary: "Dharmamitra & DharmaNexus: A New Set of Digital Tools for the Philological Study of Buddhist Texts"
- Aug 2025 Presentation, Panel "Machine Translation, Large Language Models, and Buddhist Studies" at the XXth Congress of the International Association of Buddhist Studies: "MITRA: New Research Tools for a Paradigm Shift in the Philological Study of Buddhist Texts Based on Machine Translation Technology"
- Jun 2025 Workshop Presentation, Online workshop "Navigating Indra's Net: Digital Approaches to Text Reuse-based Inter-textuality in Pre-Modern East Asian Texts" at the Hanmun Lab, Ruhr-Universität Bochum, Bochum, Germany (online): "Is training deep neural embeddings worth the effort? A preliminary investigation of different representation methods for semantic similarity tasks in Buddhist Chinese and related languages of the Buddhist tradition"
- Jun 2025 Invited Talk, DH International Workshop at Keio University, Tokyo, Japan: "From Sthiramati to Dharmamitra: Developing Digital Tools for a New Age of Philological Buddhist Studies"
- Mar 2025 Talk, CEAL (Council on East Asian Libraries) Technology Forum, Columbus, Ohio: "MITRASearch: Building Information Retrieval Systems for Classical Asian Languages in the Age of AI"

Mar 2025	Workshop, Annual Conference of the Association of Asian Studies, Columbus, Ohio: "Machine Translation for Asian Studies"
Dec 2024	Invited Talk, International Symposium "Buddhist Studies and Digital Humanities: 100 Years of the Taishō Tripiṭaka and 30 Years of SAT", Ito International Research Center, University of Tokyo, Tokyo, Japan: "MITRA Search: Exploring Buddhist Literature Preserved in Classical Asian Languages with Multilingual Approximate Search"
Nov 2024	Invited for Workshop "Case studies from current research projects - Conversations on Digital Scholarly Editing", Śivadharmā Project Headquarters, University of Naples L'Orientale, Naples, Italy: "Dharmamitra: Developing a Toolkit for Philological Work on Premodern Asian Low-Resource Languages"
Nov 2024	Invited Talk, Manuscriptology and Digital Humanities Online Lecture Series, Heidelberg University and Otani University: "Dharmamitra: New Tools for Sanskrit Translation, Grammatical Analysis, Search, and Digital Philology"
Oct 2024	Invited seminar talk, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, USA: "MITRA: Beyond Just Machine Translation for Premodern Asian Low Resource Languages"
Oct 2024	Invited Talk, Workshop "AI and the Future of Buddhist Studies", Numata Center for Buddhist Studies, University of California, Berkeley, Berkeley, USA: "Dharmamitra Search: Leveraging Multilingual Language Models for Search and Detection of Textual Reuse across Diverse Text Collections"
Aug 2024	PNC Annual Conference and Joint Meetings, Seoul, South Korea: "MITRA: Developing Language Models for Machine Translation and Search in Buddhist Source Languages"
Apr 2024	Invited Talk, Conference "International Conference on Retrospect and Prospect of Digitalization of Buddhist Resources", National Taiwan University, Taipei, Taiwan: "Massive Multilingual Machine Translation and Search for Buddhist Languages: The MITRA Project"
Mar 2024	Invited Keynote talk, "Interdisciplinary Symposium on Machine Translation and Digital Humanities", National University of Singapore, Singapore: "Dharmamitra: Enabling Massive Multilingual Machine Translation for Ancient Languages of the Buddhist Tradition"
Feb 2024	Talk, International Sanskrit Computational Linguistics Conference, Auroville, India: "Machine Translation and LLM-Powered Grammatical Explanation for Sanskrit"
Jun 2023	Invited Workshop "Machine Translation and Algorithmic Study of Classical Buddhist Texts", Dongguk University, Seoul, South Korea: "Developing Machine Translation for Ancient Buddhist Texts in Canonical Languages"
Jun 2023	Invited Talk, International symposium "Maitreya Faith/Philosophy and Buddhist Community Service", Chinese University of Hong Kong, Hong Kong: "MITRA: Developing Natural Language Processing Tools for the Languages of Buddhist Literature"

- Apr 2023 Invited Talk, Symposium "Advanced Computational Methods for Studying Buddhist Texts", Institute for the Cultural and Intellectual History of Asia, Austrian Academy of Sciences, Vienna, Austria: "Creating a Shared Semantic Vector Space for Buddhist Languages"
- Jan 2023 Talk, International symposium "Perspectives of Digital Humanities in the Field of Buddhist Studies", Universität Hamburg, Hamburg, Germany: "ChronBMM –Dating Text Corpora Using Bayesian Mixture Models"
- Jan 2023 Talk, International symposium "Perspectives of Digital Humanities in the Field of Buddhist Studies", Universität Hamburg, Hamburg, Germany: "Multilingual Semantic Mining for Text Alignment and Parallel Corpus Building for Buddhist Languages"
- Sep 2022 Talk, Conference "Deutscher Orientalistentag (DOT) 2022", Freie Universität Berlin, Berlin, Germany: "ChronBMM: Bayesian Mixture Models für die Datierung von Textkorpora"
- Sep 2022 Talk, PNC Annual Conference and Joint Meetings, University of Arizona, Tucson, USA: "The Uncertain Future of Buddhist (Machine) Translation"
- Jul 2022 Talk, 16th Seminar of the International Association for Tibetan Studies, Prague, Czech Republic: "Current State of the BuddhaNexus, a Powerful Database for Research on Buddhist Corpora"
- Jun 2022 Talk, Digital Orientalist Online Conference "Infrastructure": "Machine Translation and Buddhist Studies –First Results"
- Apr 2022 Talk, Khyentse Lecture Series, Universität Hamburg, Hamburg, Germany: "Observations on the Intertextual Relationship of the Sarvāstivāda Abhidharma Literature Preserved in Chinese Translation"
- Oct 2021 Talk, Workshop "Buddhism and Language: A Twofold Perspective: The Role of Language in Buddhist Teachings and the Role of Buddhist Sources in Linguistic Research", Ludwig-Maximilians-Universität München, Munich, Germany: "Observations on the Intertextual Relationship and Possible Authorship of the Abhidharma and Yogācāra Works Attributed to Vasubandhu"
- Sep 2018 Invited Talk, Young scholars workshop, University of Tsukuba, Tsukuba, Japan: "Towards a New Edition of the Fourth Chapter of the *Madhyāntavibhāgaṭīkā* by Sthiramati"
- Jun 2018 Invited Talk, Exzellenzcluster "Asien und Europa im globalen Kontext", Universität Heidelberg, Heidelberg, Germany: "Perspectives on Building a Multilingual Parallel Corpus of Ancient Buddhist Scriptures Based on Machine Learning"

Major Research Projects

- 2023-present **MITRA Project** – Director
Collaboration between the Center for Integrated Japanese Studies, Tohoku University, the Tsadra

Foundation, and the Berkeley AI Research Lab (BAIR), UC Berkeley
Leading the development of state-of-the-art machine translation models and semantic search functionality for Classical Asian languages
Coordinating technical infrastructure and research directions for a large team of employees, students and interns
Supervised and implemented pretraining and finetuning of large language models (LLMs) and semantic similarity search models

2018-2023

BuddhaNexus Platform – Chief Developer

Khyentse Center for Tibetan Buddhist Textual Scholarship, Universität Hamburg
Designed and implemented comprehensive database system for detecting intertextual links in Buddhist sources
Developed NLP technology for processing multilingual Buddhist texts
BuddhaNexus is now continued as DharmaNexus, part of the Dharmamitra project.

2021-2023

ChronBMM Project – Technical Lead

Heinrich-Heine-Universität Düsseldorf
Principal Investigator: Dr. Oliver Hellwig
Managed machine learning infrastructure for text dating using Bayesian Mixture Models
Implemented deep learning models for chronological analysis of historical texts
Coordinated computational resources and technical workflows

Additional relevant skills and languages

SPOKEN LANGUAGES

German (Native)
English (Fluent in speaking and writing)
Japanese (conversational in speaking, advanced in reading)
Chinese (conversational in speaking, competent in reading)

RESEARCH LANGUAGES

Sanskrit (professional in reading)
Pāli (professional in reading)
Classical Chinese (professional in reading)
Classical Tibetan (competent in reading)

Knowledge in computer science

General programming and system administration skills

- basic knowledge of C/C++, Perl, LISP/Clojure, LaTeX and Bash programming since high school; regular programming with Python since 2013

- Familiarity with Linux/Unix-type environments; administration of the web-servers of the BuddhaNexus project using docker and NGINX (buddhanexus.net)
- Managing the training of statistical models and Deep Learning models on the high performance infrastructure of the University of Duesseldorf for the ChronBMM project (Oliver Hellwig, 2021-2023); also managing the calculation of various NLP algorithms for the BuddhaNexus project on the high performance infrastructure of the University of Hamburg since 2019
- Web development: frontend and backend programming in Python (fastapi, ArangoDB) and Javascript (lit elements) for the BuddhaNexus project since 2019

Machine learning experience

- Transformer-based word segmentation tool for Sanskrit in Python with Tensor2Tensor (2018) (<https://github.com/OliverHellwig/sanskrit/tree/master/papers/2018emnlp>)
- Design and implementation of a word-vector-based system for the detection of textual reuse using nearest neighbor search (HNSW) and local alignment (Smith-Waterman) for the BuddhaNexus project based on NumPy, fastText and Faiss (2020) (<https://github.com/BuddhaNexus>)
- Dependency Parsers for Vedic Sanskrit (2021) and Latin (2022) in PyTorch (<https://github.com/sebastian-nehrdich/latin-parser>)
- Training of transformer-based language models for Sanskrit and Tibetan (SansTib 2022) (<https://github.com/sebastian-nehrdich/sanstib>)
- Since 2022, training and finetuning of language models for machine translation (Chinese to English, Tibetan to English)
- Since 2023, Training and deployment of Large Language Models (LLMs) via pytorch and deepspeed for the MITRA project with up to 10B parameters in size (<https://dharmamitra.org>; <https://huggingface.co/buddhist-nlp>)