How can User Cognition, Conversational Search, and Data Context Advance Predictive Analytics for Human-Centered AI?

Ying-Hsang Liu, Ph.D.

Professorship of Predictive Analytics (Prof. Dr. Marco Ragni), Faculty of Behavioural and Social Sciences, Chemnitz University of Technology, Germany

ying-hsang.liu@hsw.tu-chemnitz.de · ORCID 0000-0001-6504-4598 · tu-chemnitz.de/hsw/pva

Research Agenda

Predictive analytics — Modelling human behavior data to explain what happened and why, and to predict what comes next

Focus: where human cognition, information retrieval and AI meet

Four research interests

- Interactive information retrieval and knowledge organisation systems
- Human–computer interaction, including conversational agents
- Human-centered data science and AI-based design
- Cognitive insights for evaluating human-centered AI

Recent Research Projects

Recent projects linked to the professorship's areas	Information Retrieval user knowledge, eye tracking	Complex Problem Solving performance, hypotheses, strategies	Metacognition & motivation learning, distraction, personality
Airbus Conversational agents	Perceived usefulness predicts search success	Hard tasks: rephrasing, multi-step search	Users judge own performance accurately
ARDC (Australian Research Data Commons) Open research data	Data discovery behavior and search strategies	Choosing sources; judging relevance, quality	
Paradata Uppsala · Cambridge UP	Finding context for data reuse	Documenting research processes and practices	
AI & Metadata Cambridge UP volume	AI-supported metadata creation for discovery		Attitudes, skills and motivation to learn and use AI

Shared methods: mixed-effects models · eye tracking · interviews and surveys · LLM-assisted text analysis
Also at the professorship: reasoning · algorithmic thinking · human factors in autonomous driving

Airbus: Conversational Agents for Aerospace

Smart Librarian user evaluation with AIRBUS AI Research · IJHCI 2024 [1]


International Journal of Human–Computer Interaction

User Evaluation of Conversational Agents for Aerospace Domain

Liu et al. (2024), IJHCI

- **Smart Librarian:** conversational search over flight-crew documentation, evaluated in an A320/A330 cockpit simulator
- **Design:** 16 flight-school students; system (Smart Librarian vs. e-flight bag) × task difficulty; mixed-effects models
- **Finding:** perceived usefulness best predicted task score (marginal R^2 = .65) and time spent; system type alone did not
- **Implication:** satisfaction is not a reliable predictor of performance — evaluate usefulness and relevance

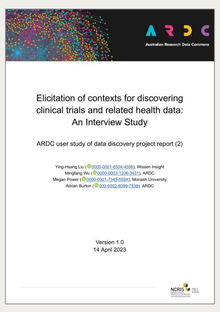
Areas: Information retrieval · Problem solving · Metacognition

ARDC: Open Research Data Discovery

Interview studies with the Australian Research Data Commons · 2022, 2023 [2]


Elicitation of Data Discovery Contexts: An Interview Study

ARDC user study of data discovery project report


Elicitation of contexts for discovering clinical trials and related health data: An Interview Study

ARDC user study of data discovery project report (2)

2022 · 24 interviews
all disciplines

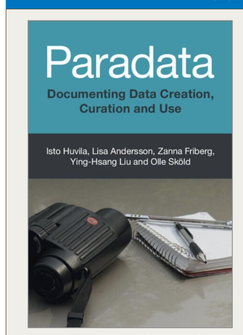
2023 · 17 interviews
clinical trials, health data

- **Aim:** how researchers discover, assess and reuse data across disciplines and in clinical trials
- **Discovery:** social networks and published sources come first; missing provenance hinders reuse
- **Clinical trials:** data quality is the most critical attribute; licensing and sharing agreements are major barriers
- **Recommended:** federated search, variable-level indexing, richer metadata and provenance

Areas: Information retrieval · Problem solving

Paradata: Data Creation, Curation and Use

Uppsala University · Cambridge University Press 2025 · ERC CAPTURE [3]


Paradata
Documenting Data Creation, Curation and Use
Isto Huvila, Lisa Andersson, Zanna Friberg, Ying-Hsang Liu and Olle Sköld

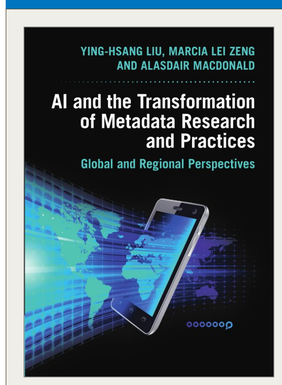
Huvila et al. (2025), Cambridge University Press

- **Paradata:** information on the practices and processes of data creation, curation and use — the context that makes data reusable
- **CAPTURE:** ERC-funded project at Uppsala University (2019–2024); the book offers the first comprehensive account of paradata
- **Two chapters (Liu & Huvila):** generating and documenting paradata (prospective, in-situ) and identifying paradata for reuse (retrospective, e.g. qualitative backtracking)
- **Open access**

Areas: Information retrieval · Problem solving

AI and Metadata: Global and Regional Perspectives

Cambridge University Press (In Press) · DCMI Education Committee [4]


YING-HSANG LIU, MARCIA LEI ZENG AND ALASDAIR MACDONALD
AI and the Transformation of Metadata Research and Practices
Global and Regional Perspectives

Liu, Zeng & MacDonald (eds.), Cambridge University Press

- **Volume:** Co-edited with Marcia Lei Zeng and Alasdair MacDonald — 18 chapters in four sections, from global surveys to ethics and governance
- **Survey:** AI in metadata practices (n = 752)
- **Finding:** Strongly positive outlook; perceived benefits are the strongest driver, lack of AI skills training the main challenge
- **Method:** Structural equation modeling (SEM); LLM-assisted extraction of topics and themes from survey responses [5]

Areas: Information retrieval · Metacognition

Connecting Cognition, Information and AI

1. **Human judgement predicts success.** Perceived usefulness and difficulty predicted search performance (marginal R^2 up to .65); gaze × task difficulty predicted answer correctness [1, 6]. *Next: link gaze with perceptions.*
2. **Context makes data reusable.** Quality, provenance, licences, metadata and paradata shape discovery and reuse, from clinical trials to the humanities [2, 3]. *Next: variable-level, federated search.*
3. **AI needs skilled, critical users.** Professionals are optimistic about AI in metadata work; training gaps and evaluating AI output are the main hurdles [4, 5]. *Next: LLM-assisted text analysis.*

References

[1] Liu, Y.-H., Arnold, A., Dupont, G., Kobus, C., Lancelot, F., Granger, G., Rouillard, Y., Duchevet, A., Imbert, J.-P., & Matton, N. (2024). User evaluation of conversational agents for aerospace domain. *International Journal of Human–Computer Interaction*, *40*(19), 5549–5568. https://doi.org/10.1080/10447318.2023.2239544

[2] Liu, Y.-H., Wu, M., Power, M., & Burton, A. (2022; 2023). Elicitation of data discovery contexts; Elicitation of contexts for discovering clinical trials and related health data: Interview studies. Australian Research Data Commons.

[3] Huvila, I., Andersson, L., Friberg, Z., Liu, Y.-H., & Sköld, O. (2025). *Paradata: Documenting data creation, curation and use*. Cambridge University Press. https://doi.org/10.1017/9781009366564

[4] Liu, Y.-H., Zeng, M. L., & MacDonald, A. (Eds.). (In Press). *AI and the transformation of metadata research and practices: Global and regional perspectives*. Cambridge University Press.

[5] Liu, Y.-H., Yang, X., & Jia, J. (2025). Assessing the effectiveness of LLMs for extracting topics and themes in survey responses. *Proc. DCMI 2025*. https://doi.org/10.23106/dcmi.952535736

[6] Liu, Y.-H., Nürnberger, A., Rettstatt, J., & Ragni, M. (2024). Saccadic eye movements and search task difficulty as basis of modelling user knowledge in information seeking. *Proc. CogSci 2024*, 46. https://escholarship.org/uc/item/3ws2g8qm

See Poster & Materials online