NA LI Ph.D.

n.li@uva.nl
https://www.linkedin.com/in/nali1001/
https://nali1001.github.io/aboutme/

INTERESTS	**Active learning	Data quality control	Data imbalance	Research asset discovery**
	*ML pipeline	Data management	Information retrieval*

RESEARCH EXPERIENCES

Active Learning for Data Quality Control | Amsterdam   2023 – 2024
- Build an AL-based data quality control framework
- Develop algorithms to solve lack of labeled data and data imbalance challenges
- Conduct experiments on large-scale ocean observatory data
- Prepare publications summarizing research findings

Research Asset Search from Heterogeneous Sources | Amsterdam   2021 - 2023
- Develop computational notebook retrieval methods using dense retrieval and LLMs
- Build evaluation dataset for retrieval method evaluation
- Develop a research asset search system including crawler, indexer and retriever
- Develop auxiliary functionalities, e.g., query reformulation and document summarization

EDUCATION

Multiscale Networked Systems, University of Amsterdam   Amsterdam, Netherlands
Ph.D. in Computer Science   2020 - 2025

School of Electronic and Information Eng., Beihang University   Beijing, China
M.S. in Information and Communication Engineering   2017 - 2020

Honors College of Beihang University   Beijing, China
B.S. in Electronic and Information Engineering   2013 - 2017

PUBLICATIONS

- **Li, Na**, Siamak Farshidi, Zhiming Zhao. "Search Multiple Types of Research Assets From Jupyter Notebook." *Software: Practice and Experience*. 2025. https://doi.org/10.1002/spe.3406

- **Li, Na**, Yiyang Qi, Chaoran Li, Zhiming Zhao. "Active Learning for Data Quality Control: A Survey." *ACM Journal of Data and Information Quality*. 2024. https://dl.acm.org/doi/10.1145/3663369

- **Li, Na**, Yiyang Qi, Ruyue Xin, et al. "Leveraging Active Learning for Ocean Data Quality Assessment: Reducing Labeling Workload and Addressing Severe Data Imbalance Challenges." *International Journal of Data Science and Analytics*. 2024. [under review]

- **Li, Na**, Yangjun Zhang, Zhiming Zhao. "A Dense Retrieval System and Evaluation Dataset for Scientific Computational Notebooks." *2023 IEEE 19th International Conference on e-Science (e-Science)*. 2023. https://ieeexplore.ieee.org/document/10254859

- **Li, Na**, Yangjun Zhang, Zhiming Zhao. "CNSVRE: A query reformulated search system with explainable summarization for virtual research environment." *Companion Proceedings of the ACM Web Conference 2023*. 2023. https://dl.acm.org/doi/abs/10.1145/3543873.3587360

- **Li, Na**, Yiyang Qi, Ruyue Xin, Zhiming Zhao. "Ocean Data Quality Assessment through Outlier Detection-enhanced Active Learning." *2023 IEEE International Conference on Big Data (BigData)*. 2023. https://ieeexplore.ieee.org/document/10386969

- Peide Zhu, **Na Li,** and Zhiming Zhao. "Retrieval-augmented Query Reformulation for Heterogeneous Research Asset Retrieval in Virtual Research Environment." *Companion Proceedings of the ACM on Web Conference 2024*. 2024.

- Siamak Farshidi, Xiaofeng Liao, **Na Li**, Doron Goldfarb, Barbara Magagna, Markus Stocker, Keith Jeffery et al. "Knowledge sharing and discovery across heterogeneous research infrastructures." *Open Research Europe* 1 (2021).

- Yuandou Wang, Spiros Koulouzis, Riccardo Bianchi, **Na Li**, Yifang Shi, Joris Timmermans, W. Daniel Kissling, and Zhiming Zhao. "Scaling notebooks as re-configurable cloud workflows." *Data Intelligence* 4, no. 2 (2022): 409-425.

- Zhao, Zhiming, Spiros Koulouzis, Riccardo Bianchi, Siamak Farshidi, Zeshun Shi, Ruyue Xin, Yuandou Wang, **Na Li**, Yifang Shi, Joris Timmermans, W. Daniel Kissling. "Notebook-as-a-VRE (NaaVRE): From private notebooks to a collaborative cloud virtual research environment." *Software: Practice and Experience* 52, no. 9 (2022): 1947-1966.

VISITINGS	**bitYoga AS** | Stavanger, Norway Sep. 2022 - Nov. 2022

bitYoga AS | Stavanger, Norway   Sep. 2022 - Nov. 2022
- Develop a testbed for Jupyter Notebook search
- Design and build a backend with a crawler and data preprocessor to collect and process large amounts of Jupyter Notebooks from various sources

Universitat Politècnica de València (UPV) | València, Spain   Oct. 2021 - Dec. 2021
- Gain first-hand experience in image processing, feature extraction, and machine learning for medical applications

Helse Stavanger HF | Stavanger, Norway (Remote)   Apr. 2021 - Jun. 2021
- Receive training on the knowledge about quantitative pathology as well as its possibilities and pitfalls at the beginner level

ACADEMIC SERVICES

Talk: BLUECLOUD seminar, Jan. 2024, Online

Oral: International Conference on Big Data, Dec., 2023, Serento, Italy
International Conference on e-Science, Oct., 2023, Limassol, Cyprus

Teaching: Teaching assistant for the DevOps course, 2021-2023

Supervision: Master thesis on intent modeling, Nov.- Mar. 2021
Master thesis on data quality control, Dec. 2022 - June 2023