

## PS4. Data Quality in Data Science

Organised by Jacob Ramlov Jensen (Go Autonomous) and Hua Lu (Roskilde University)

### Session overview

The session will focus on issues related to data quality in data science. It will feature several talks and a panel discussion, with speakers from both academia and industry. The session will cover topics such as what data quality is all about, how to achieve data quality in general, and how to improve data quality in different scenarios within data science.

### Speakers

- Kim Mouridsen, Aarhus University
- Alexander Junge, CORTI
- Nathaniel Joselson, CORTI
- Jacob Ramlov Jensen, Go Autonomous
- Hua Lu, Roskilde University
- Huan Li, Aalborg University

### Programme

13:45 - 13:50 (5 min)

- Introduction to the session (Jacob and Hua)

13:50 - 14:20 (30 min)

- What is data quality and how can we achieve it? By Kim Mouridsen (Aarhus University)

14:20 - 14:45 (25 min)

- Why do CORTI have more focus on their data than their models? By Nathaniel Joselson and Alexander Junge (CORTI)

14:45 - 15:10 (25 min)

- Human errors in data annotation and how to avoid them in an in-house AI training setup. By Nathaniel Joselson and Alexander Junge (CORTI)

15:10 - 15:30 (20 min)

- Spatial Data Quality in IoT: Issues and Mitigation Techniques. By Huan Li (AAU)

15:30 - 16:00 (30 min)

- Panel discussion (Moderator: Jacob Ramlov; panelists: Kim Mouridsen, Hua Lu, Huan Li, Alexander Junge, Nathaniel Joselson)