# BEFORE MODEL TRAINING

## CRAFTING HIGH-QUALITY TEXT ANNOTATIONS FOR MACHINE LEARNING

ALINA HERDERICH, SOFIE LABAT, JANA LASSER

A WORKSHOP FOR ICWSM 2024, JUNE 3, 2PM-6PM

**IN THIS TUTORIAL, WE WILL EXPLORE HOW TO CREATE AND APPLY HIGH-QUALITY ANNOTATION SCHEMES FOR THE AUTOMATIC CLASSIFICATION OF TEXT DATA FROM THE BOTTOM-UP.**

# JANA LASSER

PROFESSOR FOR DATA ANALYSIS
UNIVERSITY OF GRAZ

Emergent phenomena in complex social systems, humans in socio-technical environments, counterspeech, misinformation, honesty, social media algorithms

# SOFIE LABAT

PHD RESEARCHER
GHENT UNIVERSITY

Artificial emotional intelligence, emotion detection in
human interactions, conversations, coping, empathy, data perspectivism,
social biases in machine learning

# ALINA HERDERICH

PHD RESEARCHER

GRAZ UNIVERSITY OF TECHNOLOGY, HARVARD UNIVERSITY

Conceptualization and measurement of psychological constructs,
data-driven theory development, mixed-methods approaches, emotion regulation,
counterspeech, reappraisal

# TODAY'S SCHEDULE

| | | | |
|---|---|---|---|
| 2:00 - 2:15 | Alina | Presentation | Introduction |
| 2:15 - 2:35 | Jana | Presentation | How modeling choices can inform data collection and annotation |
| 2:35 - 2:55 | Alina | Presentation | A short introduction to grounded theory |
| 2:55 - 4:05 | Alina | Practical | Creating an annotation scheme using deep reading and open coding |
| 4:05 - 4:20 | | | Break |

# TODAY’S SCHEDULE

| | | | |
|---|---|---|---|
| 4:20 - 5:20 | Sofie [Alina] | Practical | Annotating data along a given framework |
| 5:20 - 5:40 | Sofie | Presentation | Trading agreement for disagreement<br>- an introduction to perspectivism |
| 5:40 - 6:00 | Alina | Discussion | Final discussion & feedback |

# DISCUSSION & REFLECTION

WHAT SURPRISED OR IMPRESSED YOU TODAY?

WHAT CHALLENGES DID YOU ENCOUNTER?

HOW DOES TODAY RELATE TO YOUR OWN PERSONAL WORK?

DID YOU FORM ANY INTENTS FOR THE FUTURE?

# WE’D BE HAPPY ABOUT YOUR FEEDBACK!

**FBR.IO/ANNOTATION**

# STAY CONNECTED!

| | ALINA | SOFIE | JANA |
|---|---|---|---|
| ✉ | alina.herderich@tugraz.at | Sofie.Labat@UGent.be | jana.lasser@uni-graz.at |
| 🌐 | hai-lina.github.io | research.flw.ugent.be/en/sofie.labat | janalasser.at |
| 𝕏 | @HaiLina16 | | @janalasser |