Ethics in AI (SSH 325/525)

Exam Date: 27/04/2025   **Time: 9:30 am-11:30 am. (Duration 2 hour)**

*Please attempt **two** questions. You need to write a brief essay (maximum 3-4 pages), that is backed by reasoning, evidence, and argued through multiple examples. You should demonstrate a keen understanding of the course's theory and reading.*

1) Value alignment is the problem of aligning human values with AI systems. What kind of problem does it represent and what kinds of questions does it raise? Critically discuss based on your readings with appropriate examples.

2) Social Robots shift values in domains where they are introduced. Do you agree with this statement? Critically discuss with appropriate examples, reason, and justification.

3) The trolley problem is an ideal representation of the kinds of ethical situations created by autonomous vehicles. These ideal situations never arise and they do not capture the complexity of real life circumstances. Discuss limitations of thinking through thought experiments while addressing AI use cases related to autonomous systems.

4) Killer robots dehumanise people and reduce them to data points. Critically discuss.