# ARTIFICIAL INTELLIGENCE

| Area | Faculty | Course name | Course instructor | Embedded ethics teacher | Informative title of intervention | Description of intervention: goal, ethical dilemmas raised, issues discussed |
|---|---|---|---|---|---|---|
| NLP | Computer Science | Introduction to Machine Learning | Dr. Yonatan Belinkov | Dr. Uri Eran | Gender bias in NLP | Background on EE, its goals and imitations; examples of gender bias in sentence completion and analogies in Hebrew and English; theoretical discussion of bias, stereotype, fairness, representational and allocation harms. |
| Machine learning | Computer Science | Advanced Topics in Computer Science 8 | Dr. Nir Rosenfeld | Dr. Avigail Ferdman | Explicit and implicit values in learning systems | Learning systems are based on conceptions of the good and are value laden. As such, they might strengthen and deepen social injustices. For example, a smart transportation app may offer more expensive trips to people to low-income people. Even well-meaning recommendation systems may inadvertently exploit human weaknesses or undermine users' autonomy. This module hones in on the implications - for users and society - on explicit and implicit values in learning systems. |
| Machine learning | Electrical & Computer Engineering | Algo and App. in Computer Vision | Prof. Lihi Zelnik | Dr. Lotem Elber-Dorozko | Ethics of computer vision technology | Machine learning bias; use and misuse of new technologies |