

# Social-minded Quality Dimensions

Tiziana Catarci

DIPARTIMENTO DI INGEGNERIA INFORMATICA  
AUTOMATICA E GESTIONALE ANTONIO RUBERTI



SAPIENZA  
UNIVERSITÀ DI ROMA

## ML, data and biases

- Machine learning systems rely on human-created data, which are the result of human-generated content, and are collected through human-created systems.
- Any bias, conscious or unconscious, in humans is embedded in algorithms and sometimes even amplified.
- Often the algorithms are based on the principle that human decision-making processes are rational, and therefore they try to find the "ratio" in historical data, but this is not always the case.
- Algorithms reproduce and augment existing inequalities or discrimination.
- Often the choices made are not understandable by the human beings.

# AI for Good

- AI ethics (and regulation): data and algorithm ethics
- Human-centred and value-oriented design of technology
- Social-minded quality measures
  - Data and algorithm fairness
  - Lack of bias
  - Diversity
- Transparency and Interpretability/Explainability

# Social minded quality measures

- Measures that evaluate the effects of algorithms on society on the basis of the degree to which they faithfully reflect the notions of ethics
- The results of an algorithm should not be influenced by attributes (protected or sensitive) that are not relevant to the activity to be performed, eg. gender, religion, age, sexual orientation, ethnicity
- **Data fairness** (bias): misuse of data and information, prejudicial behavior, favoritism, ... training data are often contaminated by the bias produced by social and historical processes → **data debiasing, fairness as a data quality measure**
  - **Individual-based fairness**: similar individuals must be treated in a similar way (where the precise notion of "similar" is defined on a case-by-case basis)
  - **Group-based fairness**: individuals are placed in groups based on the values of one or more of their protected attributes, all groups must be treated equally, eg. compared to various statistical measures such as false positives and false negatives.
  - **Fairness vs equality**; different treatment vs different impact
  - **Diversity**: ensures that all the different types of entities are considered and represented in the output of an algorithmic process, reduces ambiguity, avoids redundancy, enriches content, increases user engagement, avoids filter bubbles, echo chambers and polarization