Ethically Aligned Design for Social Robotics

Raja Chatila
Institute of Intelligent Systems and Robotics (ISIR)
Faculty of Sciences and Engineering
Sorbonne University, Paris, France

Raja.Chatila@sorbonne-universite.fr

Social Robots

- The main purpose of social robots is to interact with people and do things for people
- Social robots do not genuinely understand people and their values.
- Social robots are socio-technical objects that must be designed considering their impact on humans and society.
- Social robots may have different conflicting effects
- Design methodology that enables some alignment with human values such as privacy, intimacy, autonomy, dignity, etc., are necessary

Ethically Aligned Design

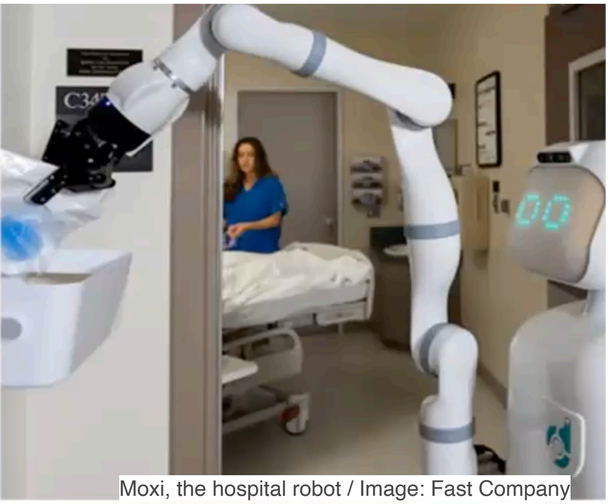
- An Ethically Aligned Design should consider system effects on all direct and indirect stakeholders, including society and the environment
- Short and long terms, local and global effects, cumulative effects
- Sustainability
- Reflection process based on ethical principles and theories
- Human values analysis and contextual prioritisation

Design Process
Questioning purpose, Involving stakeholders, Respecting values

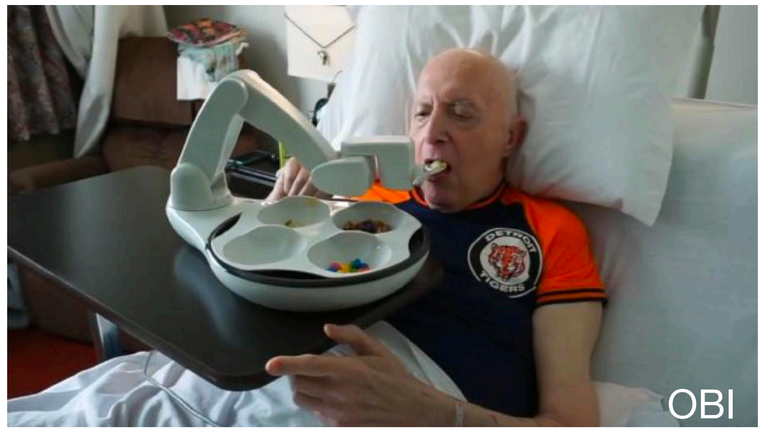
1. **Purpose and context**: What is the need to develop social robots? In which conditions and contexts? What are the benefits?
2. **Ecosystem:** Who are the stakeholders (individuals, groups, society at large, the environment)? What are their values? Are their any ethical risks?
3. **Value analysis**: Moral theories, value components, convergence and tensions
4. **Priorities** for each stakeholders during all stages (design, development, deployment, use)
5. **Technical design** solutions to comply with value priorities

Use-case: Social Robots for Care

- Why? (Usually benefits are announced)
  - Lack of care personnel
  - Reduce physical and emotional demand on care personnel
  - Specific benefits for persons (service, keeping company, information device, entertainment…)


Moxi, the hospital robot / Image: Fast Company


RIBA the robot nurse



Use-case: social robots in Elderly Homes

Stakeholders

- Elderly persons at home or in the homes
- Other persons in the homes
- Care personnel in the homes
- Families
- Care takers
- Visitors
- Technical maintenance personnel
- Robot providers
- Social security


NAO


PARO


Project ALICE

Use-case: social robots in Elderly Homes
Values

- **Dignity**
- **Human autonomy, agency, oversight (vs. infantilisation, deception)**
- **Freedom**
- **Safety, robustness**
- **Security**
- **Privacy, consent**
- **Transparency, explainability**
- **Fairness, accessibility, inclusion, cost**
- **Emotional well-being, communication, socialisation**
- **Individual, societal and environmental well-being**
- **Care personnel well-being and workload**
- **Accountability, traceability, audibility**

Robot Features

- Functions and action capabilities (behavior)
- Efficiency
- Shape and aspect (e.g., humanoid, …)
- Perception capacities
- Decision-making capacities
- Interaction, communication and human interfaces
- Ease of use
- Shared control
- Human comfort
- Reliability
- Uncertainties/impredictability due to robot behavior and human behavior

Interaction Issues

(Depend on robot complexity)

- Persistence, repeatability and engagement
- Trustworthiness : information, truth value
- Attribution of robot capabilities (e.g., sentience, emotions, anthropomorphization)
- Attachment and privation

ISO/IEC/IEEE 24748-7000:2022
Model Process for Addressing Ethical Concerns During System Design