Universität Rostock

Traditio et Innovatio

# Ethical Compliance Quantification

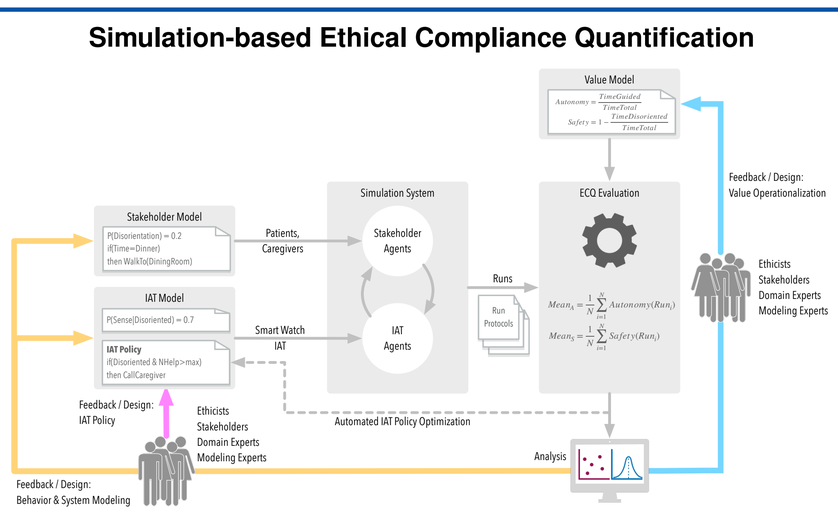
## Towards Measuring "Ethicality" of an Intelligent Assistive System

## Motivation

- Using Intelligent Assistive Technologies (IATs) poses **ethical challenges** for the stakeholders using these systems
- We need to **measure** "ethicality" of the decisions made by such IATs i.e., how does the action(s) taken by IAT impact the **ethical value set** (proposed by domain experts and stakeholders)?
- Such an analysis can help us to determine **optimal action policy** for IATs that can potentially increase their ethical-adherence

## Method

- We introduce the concept of **Ethical Compliance Quantification** (ECQ)
- ECQ evaluates the action(s) of IAT against the ethical **value model** defined by experts (e.g., Ethicists, Stakeholders, Domain Experts)
- However, performing ECQ using real-world data is **expensive** (e.g., cost, logistics) and ethically questionable (i.e., experiments with patients)
- Therefore, we perform a **simulation-based ECQ** using our simulation tool SimDEM [1] that is an agent-based modeling tool that includes **stakeholder** (i.e., PwD and caregivers) and **IAT** (i.e., smart-watch) agents
- Currently, SimDEM simulates an indoor **nursing home environment** where multiple agents interact with each other based on their role (i.e., PwD, nurse or smart-watch), abilities (e.g., a nurse agent can guide a PwD agent when disoriented) and needs (e.g., a PwD agent going for a scheduled appointment from her private room)




## Value Model

Depending on the requirements and domain, various ethical values can be incorporated into the value model for ECQ. The method to measure such value set depends upon the experts and stakeholders. We illustrate **two approaches** ($Type_I$ & $Type_{II}$) to measure **violation** of each value: **Autonomy** and **Safety**, as a proof-of-concept.

- ***Autonomy$_I$*** Violation of $Autonomy_I$ ($\neg A_I$) is defined as the percentage of the time PwD agent is guided ($TimeGuided$) by the nurse agent per simulation
- ***Safety$_I$***: Violation of $Safety_I$ ($\neg S_I$) is the number of instance where PwD agent is $Disoriented$ and there is no help provided ($NoIntervention$) i.e., neither nurse guidance nor smart watch intervention
- ***Autonomy$_{II}$***: Violation of $Autonomy_{II}$ ($\neg A_{II}$) is defined as the sum of number of smart-watch intervention ($N_{SI}$: technical-intervention) and guided episodes ($N_{GE}$: human-intervention) per simulation
- ***Safety$_{II}$***: Violation of $Safety_{II}$ ($\neg S_{II}$) is defined as the time taken by a PwD agent to regain the orientation ($TimeReorientation$) after being disoriented for each disorientation episode ($i$) averaged over the total number of episodes ($n$) per simulation

$$\neg A_I = \frac{TimeGuided}{TotalTime} \qquad \neg S_I = \sum^{n} [Disoriented_k \wedge NoIntervention_k] = \begin{cases} 1, & \text{if True} \\ 0, & \text{otherwise} \end{cases}$$

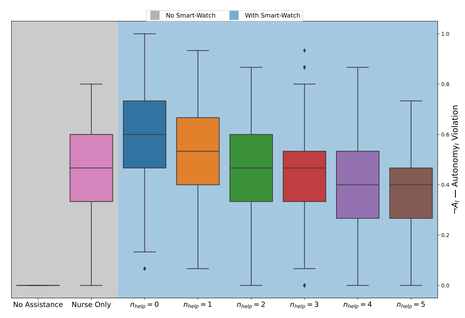
$$\neg A_{II} = N_{SI} + N_{GE} \qquad \neg S_{II} = \frac{1}{n}\sum_{i=1}^{n} (TimeReorientation)_i$$

## Results

Results of ECQ for violation of **Autonomy** and **Safety** are presented against the smart-watch parameter $n_{help}$ which represents different assistive strategies along with the results without using a smart-watch. Here, $n_{help}$ Is the **number of failed navigation interventions before a nurse agent is called** and represents how soon a caregiver intervention is called. In all figures shown bellow, **less is better**.

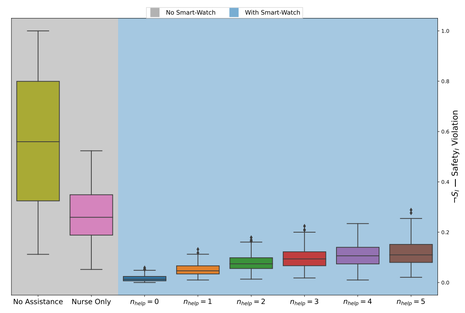
### $Autonomy_I$ Violations versus Assistance

- Without a smart watch, PwD agents stay **more autonomous** (i.e., $\neg A_I$ decreases) as the nurse agent is mostly **unaware** if PwD agents are disoriented or not unless they visually identify a PwD agent by chance
- Generally, as we increase $n_{help}$, autonomy improves because smart watch tries (more frequently) to reorient the PwD agents before a caregiver intervention is called



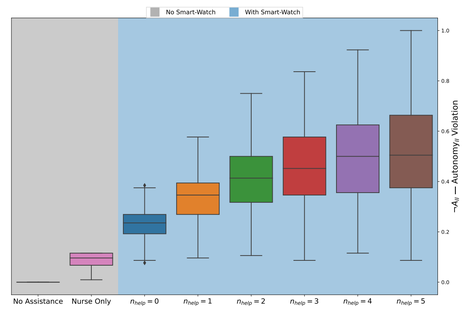

### $Safety_I$ Violations versus Assistance

- More safety violations ($\neg S_I$) occur without smart watch as nurse agents are unaware of the disoriented PwD agents
- Smart watch **significantly improves** safety as the system can help PwD more efficiently
- Increasing $n_{help}$ worsens safety because the nurse intervention gets delayed



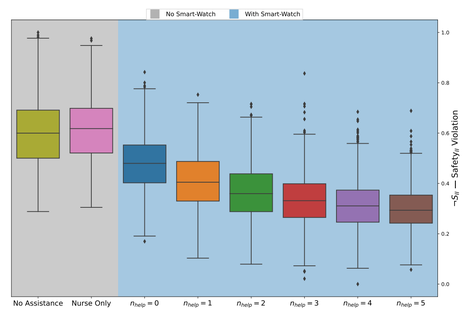

### $Autonomy_{II}$ Violations versus Assistance

- PwD agents become less autonomous as we use a smart-watch and increase $n_{help}$ because increasing $n_{help}$ increases technical interventions i.e., increase in smart watch interventions ($N_{SI}$)




### $Safety_{II}$ Violations versus Assistance

- Addition of the smart-watch and increase in $n_{help}$ **improves safety** because the time taken by PwD agents to get reoriented ($TimeReorientation$) decreases as the smart-watch is allowed to intervene more frequently and gives PwD agents more time to recover




- The results reported here using the ECQ provide interesting insights
- However, main outcome of this work is not the exact results and value model
- But, rather, how such a simulation-based approach can be setup to analyse and model assistive strategies based on expert-defined ethical value model

## References


[1] Shaukat, M.S., Hiller, B.C., Bader, S., Kirste, T., 2021. SimDem A Multi-agent Simulation Environment to Model Persons with Dementia and their Assistance, in: 4th International Workshop ARIAL held at IJCAI 2021.



**M. S Shaukat** [a], **J.-C. Põder** [b], **Sebastian Bader** [a], **Thomas Kirste** [a]

[a] COMPUTER SCIENCE DEPARTMENT | MMIS | **Universität Rostock**, Germany
[b] FACULTY OF THEOLOGY | **Universität Rostock**, Germany