FRTF25: Ethics

Social Robots for the Disabled

Rouyi Zhao
Machine Learning, Systems & Control
LTH, Lund University
Lund, Sweden

Basim Elessawi
Machine Learning, Systems & Control
LTH, Lund University
Lund, Sweden

Jingmo Bai
Machine Learning, Systems & Control
LTH, Lund University
Lund, Sweden

I. Introduction (*Heading 1*)

Social robotics is a new up and coming field in the present time automation industry. With global market forecasts predicting a compound annual growth rate of over 34.3% through the upcoming 5 years [19], predominantly in healthcare and education [20]. However, the stigma against the integration of AI into societal roles (such as caretakers) still remain prevalent, with social cues, such as disappearance of jobs with the emerging AI sector weighing against an assimilation of AI into several industrial and social positions [15-17], despite being proven economically untrue [18].

The importance of AI however is not substitutive in nature, but more complementary; taking the examples of nursing homes, healthcare, and domestic care all of which are fields where social robots, interactive both physically and mentally, are slowly being implemented via closely monitored trials and assessments [ref]. The number of elderly people, or senior citizens, over the age of 60 are on a constant rise, In 2020 elderly aged over 60 outnumbered children below the age of 5[ref], and according to the WHO, the number is only expected to increase in the upcoming decades from 900 million in 2015 to about 2 billion in 2050 [2], in comparison the growth rate for nurses in the past half decade has been about 0.7% in the US [3], and is projected to increase about 6% over the next decade [4], with even more about to retire in the upcoming years due to a median age of 52 [5].

Ignoring the obtrusive facts mentioned in the statistics, according to the GMC (General Medical Council) a disabled person is described as someone who has a “physical or mental impairment that has a substantial and long-term adverse effect on the person's ability to carry out normal day-to-day activities.” [1]. This includes, but is not limited to, elderly, children with disabilities, and people disabled due to physical or mental trauma. Putting aside the increasing number of elderlies, current situations towards senior citizens with disabilities are not particularly favorable, neglect and abuse to the elderly are prominent in nursing homes, and even from domestic caretakers (family members), according to the WHO, around 1 in 6 people over the age of 60 with disabilities go through some form of abuse or neglect, with 64.2% of caretakes reporting having committed some form of abuse. Negligence can go a long way to a disabled person, from inducing mental strain to deprivation of medication and food, leading to in extreme cases, death [2]. Apart from people aged over 60, according to UNICEF, there are about 240 million children, roughly 10% of all children, living with disabilities [6], of which 32% experience violence, neglect, abuse, and are bullied by their peers [7]. In order to prevent psychological and physical torture, an efficient and regulated system is needed.

The caregiver crisis arises from the underpayment of nursing staff, along with the rapidly expanding population of elderly and disabled. Informal caretakers, most commonly family members, are hard-pressed to provide care whilst fulfilling the needs of the disabled co-habitants, it was recorded that an approximate cost of $62,000 a year in extra expenses is used when taking care of a loved one [13], along with their own individual social roles, leads to significant neglect, both as a physical and mental entity, and being absent during times of need (medicine, treatment etc.) [14].

Social robotics, when used in domestic care help monitor patients whilst at the same time interact with them in accordance with the social role they occupy. Along with the presence of a supervisory figure, they are able to provide significant assistance for caretakers and their targeted care receivers effectively alleviating the caretaker crisis. Social robots have been reported to be effective in settings where both physically and mentally disabled individuals are involved, including settings of caretaking, monitoring, and rehabilitation [9-12]. However, the advent of AI has not been completely harmless, from insignificant accidents such as displaying the wrong ads, to accidents in defense departments which could potentially trigger wars, or even set back the economy decades [21]. As such it is extremely critical to handle the matter in a socially and ethically responsible manner. This report discusses the ethical approach to social robots which are being used or proposed to be used in order to take care of the disabled, as well as examining present time implementations [8].

II. Aspects and their significance

A. Transparency

As one of the fundamental properties in AI governance, it's something that can present a solution for underlying ethical questions[27]. Since many of the present-day AI systems are still referred to as a "black-box", which means, the specific workings of these AI technology are not disclosed by the relevant authorities[27]. With such a lack of openness or "Transparency" might lead to a range of related problems and can even cause seriously negative effects. Therefore, against this background of conditions, transparency considerations are needed for ensuring the individual's rights to know that how the critical decisions are made by AI algorithm or the accountability for mistakes can be more fairly traced, so that the benefits of its development can be made accessible for all stakeholders.

Even without the existence of the black-box, many data-driven technologies have extremely complex internal networks, networks with hundreds of even thousands of nodes with tree-digit hidden layers, are exceptionally hard to explain, and with increasing complexity the relationship between the inputs and outputs becomes increasingly blurred. Sufficient transparency requires breaking the barriers of opacity as mentioned by Högberg and Larsson: Opacity due to confidentiality, Opacity due to technical illiteracy, and opacity due to scale of operation of algorithms [28].

To achieve an ethical level of transparency stakeholders and those developers will need to provide relevant information on how decisions are made, whilst also mentioning what data is being used, similar to the model suggested by Deloitte, the 'Glassbox', as opposed to the black box [29]. Transparency is at an optimal when information pertaining to, or that could effect, the end-user is known to them whilst also protecting them from any malicious threats which could result from the compromised information.

B. Accountability

The term accountability is very often misinterpreted as the word 'blame', in AI and data-driven methods, accountability refers to responsibility upheld by each individual along the development and usage chain.

When it comes to accountability, there are three main elements which are auditability, risk management, and redress.

In the role of developers and companies of this kind of neurorehabilitation robot, they should ensure their products or services are auditable, for example, all the development process, algorithms, source of training data, and the log data of the AI systems, both before and after their development, deployment, and use[30]. If some kind of negative impact occurs, those data and algorithms should be open to be audited. And also it is indispensable that there are mechanisms established to oversee the responsibility and accountability of AI products and services.

The second part is risk management, the developers and companies should carry out a risk or impact assessment of the AI products during the development process and inform the consumers or any other stakeholders of potential vulnerabilities or risks[30]. For example, if using this kind of robot could contribute to any kind of injury physically or mentally, even if the possibility of injury happens is not known or not tested, the doctors and the patients should be informed of the risk. To minimize the negative impacts, developers and companies need to have the ability to respond to conceivable outcomes like injury and damage to humans and property, and also to report on the causes that contribute to such kinds of outcomes[30].

All stakeholders along the chain from product development to use, like health systems, insurers, developers, and regulators should bear enough liability to ensure AI safety. Insurers can pay for the compensation to the victims of an AI injury. However, cure isn't a substitute for prevention, most immature mistakes should be screened heavily by regulators to prevent any risky AI products that could be detrimental to vulnerable end users, especially the disabled, elderly, and children.

Due to an absence of fixed set of standards and regulations for accountability, any additional major AI incidents should be litigated in special courts with expertise in adjudicating AI cases to determine how should responsibilities be distributed.

C. Fairness

Every AI model has some inherent bias encoded within the model; this bias comes from the fact that there is bias saturated into every part of society. The data used to train AI models contain historical bias, whilst training the models validation and test data sets include their own algorithmic and validation bias, whilst the developer himself induces some behavioral, or social bias.

This sort of bias is displayed in various ways, such as unfair allocation of resources and welfare based on nationality, gender, and disability [31], repetition of social stereotypes like White male CEO's or female nurses.

Unlike transparency and accountability, fairness is much more inherent, and thus difficult to eliminate. As a solution multiple toolboxes targeting fairness in ML or data-driven technologies have been created, such as python's AI Fairness 360 library, Fairlearn toolbox, Google's What-If Tool etc.

However, absolute fairness or completely generalized models are not necessarily considered positive in all situations, in the case of social robots for example, whilst dealing with mentally sensitive patients, a sentience which can relate culturally would tend to be more attractive towards end users as compared to a completely neutral model [26]. Thus instead of de-biasing AI models, most fairness tools tend towards eliminating discriminations which may cause harm to end - users, or to ensure the bias doesn't negatively affect the output.

III. Case-in-points

A. DRESS-eAI

DRESS-eAI is a data-driven risk assessment method proposed by Felländer et al. aimed at developing an ethical AI solution for AI risk management and governance in companies where AI is implemented in the framework of their routine processes. This was done by first identifying various ethical

and societal risks that may occur as a result of using AI and aims to ensure their prevention. Eight core risks were identified: Privacy intrusion, amplified discrimination, violation of autonomy and independent decision making, social exclusion and segregation, harm to safety, harm to security of information, misinformation and disinformation, and prevention of access to public service [23].

The framework of DRESS-eAI and similar governance frameworks are critical in the integration of AI based models into social and industrial roles. The report also discusses possible pitfalls due to data and fairness issues which could be avoided in advance, whilst also proposing 6 phase model of risk assessment which addresses a large portion of the dubiousness of AI ethics.

B. CARESSESS

The CARESSESS project was established in order to develop SARs which are culturally sensitive, the CARESSES trial was then used to test how culturally competent robots helped improve health standards for senior citizens in old care homes. It was tested on a specimen of 45 senior citizens aged above 65 from several old homes, the individuals had identities belonging to English, Japanese, and Indian descent [26]. The methodology of the experiment was to pair the participants with a robot from either the experimental group, which was more culturally robust, or the control group, which were still interactive robots, but not as culturally flexible, a third and final group called 'Care as usual' were without any robots. The results showed that both the control groups and experimental groups had improvements to their mental wellbeing and satisfaction factors, whilst the 'care as usual' had the opposite reaction, and the experimental group showed higher improvement than the control group, by a factor of 8 [24].

The CARESSESS project is a prime example of the difference between absolute fairness and adjustable bias, whilst the patients in the model showed improvements in both models the model designed to be more culturally robust had outperformed the neutral model by far, of course this could be due to the inherent racial bias, but due to it being a controlled and monitored trial, the effects would not have been apparent.

In terms of transparency, the technological details were not mentioned in their report, however they mentioned handing out leaflets to the residents explaining the process in detail, whether or not this included the technical aspect was not made clear. However, the data required and recorded was known to the end-users and their informal caretakers. Consent forms were also signed in order to reduce any accountability issues that may arise as a result of the trial [26].

C. Neurorehabilitation

The Injuries to the central nervous system can result in a lack of functionality of motor neurons, this lack of functionality results in the muscles not receiving commands from the brain, and thus results in either partial or complete lack of mobility from said muscles or limbs. Neurorehabilitation is the process by which a disabled individual can recover partially or completely with the assistance of technology and therapy, some of which include invasive brain stimulation [22]. Socially assistive robots (SAR) can be used to provide neurorehabilitation in place of therapists, and even more so be able to do it at a remote, decentralized access point, such as the individuals home, as opposed to the regular therapy sessions and long hospital visits [25].

Neurorehabilitation being originated from a clinical therapeutical method, shares many similarities to how social robots interact with individuals with inherent disabilities, such as kids with autism. Transparency trials in similar cases had shown significant ease in reporting relevant data and information [10]. Rakhymbayeva et al. in their research had provided clinical trial data, as well as demonstrated and explained their end-user interaction with the robot, what to expect and how their trail system worked to both parents and kids, accountability was also settled using consent forms [10]. Similarly the transparency in Neurorehabilitation case should include:

Documentations of the process of collecting and labelling the database, as well as the used algorithms, make a clear vision of the foundation for every decision made in the AI system. In this case which is, for instance, what kind of signals are detected by the sensor before it reacts to the current behaviors. According to this evidence, a fairer and more observable result will be obtained when the issue of accountability is clarified.

Furthermore, the patients should be explained by the doctors and the creator organization about how the machine can help or even replace them in performing physical movements that they cannot perform on their own and how exactly does this robot make executive decisions based on the weak bio-signals provided by the patient. In order to ensure to avoid the risk of improper handling of patients in unknown situations.

Moreover, when such neurorehabilitation robots change from a hospital-centered model to a home-centered model of neurorehabilitation[ref5], the movement data and any other information collected by the sensors on the robotics, should also be informed to the users by the creator organization. It's not only the privacy issues, but also about the security issues of user's personal information, there is potential possibilities that the information could be intercepted by cyber criminals or hackers during the transmission process, which might cause serious property and even personal safety risks to the user

IV. CONCLUSIONS

Social robots being used in domestic care situations seems to be a suitable solution to the caregiver crisis. However, without the correct guidelines, standards, and governance AI can prove to be a disaster beyond the norm, the recent event of an AI chatbot being tested on fake profiles had amongst one of the 'people', who was profiled as suicidal, advised it to kill themselves, the result was later addressed as 'inconsistency'. Albeit being a fake patient, events like this are possible to

occur in reality, and when dealing with mentally sensitive disabled people, it could lead to many lives being lost, a mentally unstable young person may also be lead further into dementia and hallucinations causing unpredictable chains of events. Although this is all speculation, it is nevertheless a possibility. Thus, a very stringent method of governance and screening is required, whilst test trials are being used now in order to prevent any unprecedented situations, SAR's should be monitored and updated regularly by an authoritative figure, who can take responsibility and accountability for the robot. Innovation is growing at an alarming pace, testing methods are not able to keep up with development, as engineers the effects of any developments in the practical world needs to be taken only after complete assurance is acquired, in the interest of the safety and wellbeing of humanity.

REFERENCES

[1] *Education standards, guidance and curricula - Gmc*. (n.d.). Retrieved November 24, 2022, from https://www.gmc-uk.org/education/standards-guidance-and-curricula

[2] World Health Organization. (n.d.). *Abuse of older people*. World Health Organization. Retrieved November 24, 2022, from https://www.who.int/news-room/fact-sheets/detail/abuse-of-older-people

[3] *Industry market research, reports, and Statistics*. IBISWorld. (n.d.). Retrieved November 24, 2022, from https://www.ibisworld.com/industry-statistics/market-size/nursing-care-facilities-united-states/#:~:text=US%20in%202022%3F-,The%20market%20size%20of%20the%20Nursing%20Care%20Facilities%20industry,to%20increase%204%25%20in%202022.

[4] Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, Registered Nurses, at https://www.bls.gov/ooh/healthcare/registered-nurses.htm

[5] *Nurses in the Workforce*. ANA. (2017, October 14). Retrieved November 24, 2022, from https://www.nursingworld.org/practice-policy/workforce/

[6] Hosta, S. (n.d.). *Fact sheet: The World's nearly 240 million children living with disabilities are being denied basic rights*. UNICEF. Retrieved November 24, 2022, from https://www.unicef.org/turkiye/en/press-releases/fact-sheet-worlds-nearly-240-million-children-living-disabilities-are-being-denied

[7] Bawden, A. (2022, March 17). *Almost a third of disabled children and teenagers face abuse, global study finds*. The Guardian. Retrieved November 24, 2022, from https://www.theguardian.com/society/2022/mar/17/almost-a-third-of-disabled-children-and-teenagers-face-abuse-global-study-finds#:~:text=About%20one%2Dthird%20of%20young,violence%2C%20according%20to%20international%20research.

[8] Laban, G. (2021, May 11). Caregivers' little helpers: How can social robots support informal caregivers with monitoring the health and well-being of care recipients? ENTWINE. Retrieved November 24, 2022, from https://entwine-itn.eu/caregivers-little-helpers-how-can-social-robots-support-informal-caregivers-with-monitoring-the-health-and-well-being-of-care-recipients/

[9] Valentí Soler, M., Agüera-Ortiz, L., Olazarán Rodríguez, J., Mendoza Rebolledo, C., Pérez Muñoz, A., Rodríguez Pérez, I., Osa Ruiz, E., Barrios Sánchez, A., Herrero Cano, V., Carrasco Chillón, L., Felipe Ruiz, S., López Alvarez, J., León Salas, B., Cañas Plaza, J. M., Martín Rico, F., Abella Dago, G., & Martínez Martín, P. (2015). Social Robots in advanced dementia. *Frontiers in Aging Neuroscience*, *7*. https://doi.org/10.3389/fnagi.2015.00133

[10] Rakhymbayeva, N., Amirova, A., & Sandygulova, A. (2021). A long-term engagement with a social robot for autism therapy. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.669972

[11] Gilbert, C. (n.d.). *How social robotics can benefit disabled people*. AbilityNet. Retrieved November 24, 2022, from https://abilitynet.org.uk/news-blogs/how-breakthroughs-social-robotics-are-set-benefit-disabled-and-neurodiverse-people

[12] Kubota, A., & Riek, L. D. (2022). Methods for robot behavior adaptation for Cognitive Neurorehabilitation. *Annual Review of Control, Robotics, and Autonomous Systems*, *5*(1), 109–135. https://doi.org/10.1146/annurev-control-042920-093225

[13] *The caregiving crisis*. The Center for Workforce Inclusion. (2021, October 12). Retrieved November 24, 2022, from https://www.centerforworkforceinclusion.org/the-caregiving-crisis/

[14] Bermanstaff. (2021, April 14). *Abuse of disabled adults: Data & Statistics, how to recognize & report: Berman & Riedel, LLP*. Berman Lawyers. Retrieved November 24, 2022, from https://bermanlawyers.com/neglect-abuse-disabled-adults/

[15] *Biggest Artificial Intelligence controversies: Racism, sexism and 'becoming sentient'*. The Indian Express. (2022, June 15). Retrieved November 24, 2022, from https://indianexpress.com/article/technology/tech-news-technology/biggest-artificial-intelligence-controversies-racism-sexism-and-becoming-sentient-7969270/

[16] Tomar, D. (2022, November 22). *Controversial topic: Artificial intelligence*. Academic Influence. Retrieved November 24, 2022, from https://academicinfluence.com/inflection/controversial-topics/controversial-topic-artificial-intelligence

[17] *Robots and ai taking over jobs: What to know about the future of Jobs*. Built In. (n.d.). Retrieved November 24, 2022, from https://builtin.com/artificial-intelligence/ai-replacing-jobs-creating-jobs

[18] *Impact of AI on future job roles*. Times of India Blog. (2022, November 19). Retrieved November 24, 2022, from https://timesofindia.indiatimes.com/blogs/voices/impact-of-ai-on-future-job-roles/

[19] *Social robots market size, share, growth, trends: 2022 - 2027*. Social Robots Market Size, Share, Growth, Trends | 2022 - 2027. (n.d.). Retrieved November 24, 2022, from https://www.mordorintelligence.com/industry-reports/social-robots-market

[20] *Social Robots Market: Global Industry Analysis and Forecast (2021-2027) by end user, application, industry*. MAXIMIZE MARKET RESEARCH. (2022, February 22). Retrieved November 24, 2022, from https://www.maximizemarketresearch.com/market-report/global-social-robots-market/78802/

[21] Arnold, Z., & Toner, H. (2021). Ai accidents: An emerging threat. https://doi.org/10.51593/20200072

[22] Semprini, M., Laffranchi, M., Sanguineti, V., Avanzino, L., De Icco, R., De Michieli, L., & Chiappalone, M. (2018). Technological approaches for neurorehabilitation: From robotic devices to brain stimulation and beyond. *Frontiers in Neurology*, *9*. https://doi.org/10.3389/fneur.2018.00212

[23] Felländer, A., Rebane, J., Larsson, S., Wiggberg, M., & Heintz, F. (2022). Achieving a data-driven risk assessment methodology for ethical AI. *Digital Society*, *1*(2). https://doi.org/10.1007/s44206-022-00016-0

[24] Papadopoulos, C., Castro, N., Nigath, A., Davidson, R., Faulkes, N., Menicatti, R., Khaliq, A. A., Recchiuto, C., Battistuzzi, L., Randhawa, G., Merton, L., Kanoria, S., Chong, N.-Y., Kamide, H., Hewson, D., & Sgorbissa, A. (2021). The caresses randomised controlled trial: Exploring the health-related impact of culturally competent artificial intelligence embedded into socially assistive robots and tested in older adult care homes. *International Journal of Social Robotics*, *14*(1), 245–256. https://doi.org/10.1007/s12369-021-00781-x

[25] Lambercy, O., Lehner, R., Chua, K., Wee, S. K., Rajeswaran, D. K., Kuah, C. W., Ang, W. T., Liang, P., Campolo, D., Hussain, A., Aguirre-Ollinger, G., Guan, C., Kanzler, C. M., Wenderoth, N., & Gassert, R. (2021). Neurorehabilitation from a distance: Can intelligent technology support decentralized access to quality therapy? *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.612415

[26] Papadopoulos, C., Hill, T., Battistuzzi, L., Castro, N., Nigath, A., Randhawa, G., Merton, L., Kanoria, S., Kamide, H., Chong, N.-Y., Hewson, D., Davidson, R., & Sgorbissa, A. (2020). The caresses study

protocol: Testing and evaluating culturally competent socially assistive robots among older adults residing in long term care homes through a controlled experimental trial. *Archives of Public Health*, *78*(1). https://doi.org/10.1186/s13690-020-00409-y

[27] What is transparency? - ethics of ai. (n.d.). Retrieved November 22, 2022, from https://ethics-of-ai.mooc.fi/chapter-4/2-what-is-transparency

[28] Högberg, C., & Larsson, S. (2022). AI and Patients' Rights: Transparency and information flows as situated principles in public health care. I K. de Vries, & M. Dahlberg (Red.), De Lege – Yearbook Uppsala Faculty of Law 2021: Law, AI & Digitalization (s. 401-429). (De Lege; Vol. 2021). Iustus förlag. [29]

[29] *Transparency and responsibility in Artificial Intelligence - Deloitte*. (n.d.). Retrieved November 24, 2022, from https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/innovatie/deloitte-nl-innovation-bringing-transparency-and-ethics-into-ai.pdf

[30] Smuha, Nathalie A. "The EU approach to ethics guidelines for trustworthy artificial intelligence." Computer Law Review International 20.4 (2019): 97-106.

[31] Guardian News and Media. (2021, November 21). *DWP urged to reveal algorithm that 'targets' disabled for benefit fraud*. The Guardian. Retrieved November 24, 2022, from https://www.theguardian.com/society/2021/nov/21/dwp-urged-to-reveal-algorithm-that-targets-disabled-for-benefit