

Requisitos de Confiança para Robôs Socialmente Assistivos

Larissa Rodrigues da Costa¹, Jaelson Castro¹

Centro de Informática - Universidade Federal de Pernambuco, Recife/PE, 50740-560,
Brasil Email: {lrc, jbc}@cin.ufpe.br

Resumo Contexto: No campo de pesquisa dos Robôs Socialmente Assistivos (SARs) o estudo da Interação Humano-Robô (HRI) tem o importante papel de indagar a forma como o aspecto humano e robótico colaboram entre si e qual o espaço do robô na vida humana. Alguns aspectos vêm sendo bastante estudados na literatura, dentre eles a confiança, que é crucial e bastante complexa, composta de diversos fatores que precisam de algumas condições necessárias para ser atingida, condições essas, que a Engenharia de Requisitos pode contribuir para compreender, moldar e analisar quais requisitos são fundamentais melhorando assim a aceitação da utilização desses robôs. **Objetivo:** Essa pesquisa propõe a construção de um catálogo de requisitos não funcionais chamado NFR4TRUST, que possa ajudar a aumentar a confiança humana na utilização de robôs no domínio de SARs. **Método:** Inicialmente, será realizado um levantamento bibliográfico visando verificar o estado da arte em HRI no contexto de confiança, segurança e taxonomias existentes. A partir deste estudo, será construída uma proposta de taxonomia que auxiliará a descoberta de requisitos para SARs. A técnica STPA será utilizada como método para avaliar as ameaças a confiança, enquanto a validação do catálogo proposto será realizada por especialistas e elaboração de um estudo de caso. **Resultados esperados:** Com este trabalho criaremos o Catálogo NFR4TRUST que permitirá apoiar as etapas de elicitação e especificação de requisitos referentes a confiança no domínio proposto. **Conclusão:** Os resultados deste trabalho permitirão incrementar, avaliar e melhorar a confiança humana na utilização dos SARs.

Keywords: Engenharia de Requisitos · Interação Humano-Robô · Robótica Socialmente Assistiva · Confiança · STPA .

Nível: Mestrado

Ano de Ingresso no Programa: 2019.2

Época prevista de conclusão: 2021.2

1 Introdução e Caracterização do Problema

Os Robôs Socialmente Assistivos (*Socially Assistive Robotics* - SARs) são robôs que prestam assistência ao usuário humano, com ênfase na interação social. Eles

vêm sendo objeto de interesse, uma vez que é crescente a construção de ativos tecnológicos que visam melhorar a qualidade de vida de pessoas com algum tipo de limitação, física ou mental [1].

O campo da Interação Homem-Robô (*Human-Robot Interaction* - HRI) que é considerada por Yan et al. (2014) [2] o coração de um robô social, vem buscando compreender e moldar como se comporta a colaboração entre o aspecto humano e o robótico, principalmente no que se refere a confiança e segurança humana, que são elementos-chaves no desenvolvimento de relacionamentos eficazes [3,4]. Portanto, inúmeros pesquisas na área de HRI tentam identificar fatores que ajudam a promover a confiança humana em relação aos robôs [5].

O conceito de confiança se estende a muitas disciplinas e pode ser abordado sob várias perspectivas e assim como é difícil definir um significado fixo acerca de tal tema, é também uma tarefa árdua estabelecer métricas nas quais pode-se medi-la. Portanto, neste trabalho, a definição de confiança será construída delimitando os conceitos referentes a grande área da robótica para chegar a uma definição de confiança própria aos SARs. Logo, será necessário selecionar os fatores que constituem a forma de avaliar esse conceito de Confiança e organizá-los através da utilização da taxonomia para melhor compreender como se dividem e se relacionam, e assim poder avaliá-los.

Outro aspecto a ser considerado é a relação entre segurança e confiança, haja visto que quanto menor o nível de segurança percebido, menor a chance do indivíduo confiar na utilização do ativo robótico [6]. Sendo assim, a taxonomia de Confiança deveria incluir aspectos de segurança para tornar-se mais completa e adequada.

As atividades e o processo da Engenharia de Requisitos são essenciais no desenvolvimento de diversos sistemas, uma vez que ela busca evitar a introdução de defeitos, mal entendidos entre pessoas envolvidas no processo de desenvolvimento, como também, busca minimizar substancialmente o risco de fracasso desses sistemas. Contudo, é possível observar que em muitas áreas geralmente se inicia o desenvolvimento de um determinado sistema antes de capturar seus requisitos.

Nos sistemas robóticos, os requisitos têm um papel fundamental para garantir especificações, descrições de como o sistema deveria se comportar e restrições sobre a operação dele. Existem alguns padrões internacionais de segurança que servem como base para garantir alguns requisitos de segurança, como os elaborados na ISO 13482:2014, que se concentra nos requisitos para robôs de cuidados pessoais em aplicações não-médicas. Porém, não existe ainda um padrão específico para segurança de SARs, logo, se torna mais difícil avaliar quais são os requisitos essenciais e exclusivos desse domínio, e quais deles tem maior impacto na confiança. Requisitos como os Não Funcionais (*Non Functional Requirements* - NFRs) que são atributos de qualidade e que estão diretamente relacionados à funcionalidade de um sistema, têm um papel relevante durante o desenvolvimento de um robô, pois se tais requisitos não são levados em consideração, então o sistema poderá ser inconsistente e de baixa qualidade. É de grande valia investir na elaboração de tais requisitos nessas aplicações que requerem intera-

ção e colaborações próximas entre humanos e robôs, bem como contato físico, influência social e psicológica, para assegurar uma consistência no conjunto de atributos que o sistema robótico deveria suportar para ser considerado confiável.

O conhecimento sobre os Requisitos Não Funcionais pode ser organizado em forma de catálogo, facilitando o compartilhamento, gerenciamento e melhoramento de informações [7]. O propósito deste trabalho é criar um Catálogo de Requisitos Não Funcionais chamado NFR4TRUST voltado para a Confiança em Robôs Socialmente Assistivos que influencie a aceitação e utilização deste tipo de robô. Esse Catálogo será útil tanto na fase inicial de desenvolvimento e projeto, como auxiliará na avaliação de robôs já desenvolvidos.

2 Fundamentação Teórica

Existem diversas categorias de robôs dentro da literatura e muitas vezes a forma como são classificadas variam de acordo com cada pesquisador ou domínio de pesquisa. Por exemplo, a Robótica Socialmente Assistiva (SARs) é uma das três grandes áreas da robótica classificada no campo de estudo da Interação Humano-Robô (HRI). De acordo com Feil-Seifer et al. (2005) [8] ela é a intersecção entre as outras duas áreas que são: A Robótica Assistiva (AR), da qual os robôs provêm assistência às pessoas com alguma deficiência física, possibilitando-os melhoria em sua independência; e a Socialmente Interativa (SIR), que possui alguma forma de interação social com o usuário humano para execução de sua tarefa principal. Em SARs o objetivo é criar uma interação próxima e eficaz com um usuário humano com o propósito de dar assistência e alcançar um progresso mensurável na recuperação, reabilitação, aprendizagem, entre outros [8].

A HRI busca projetar e avaliar sistemas robóticos que serão utilizados para uso por ou com humanos e cujo o contato com os usuários sejam eficientes e aceitáveis. Como é uma área altamente interdisciplinar, algumas dificuldades podem ser percebidas no que se refere a pluralidade de vocabulários, metodologias e métricas utilizadas pela comunidade [9]. Existem alguns conjuntos de problemas e desafios que geralmente geram esforços de pesquisa da comunidade de HRI, dentre eles o interesse de identificar os fatores que promovem a confiança humana e a segurança em relação aos robôs.

A confiança desempenha um papel importante na interação dos seres humanos; a eficácia da comunicação, o aprendizado e a resolução de problemas são quesitos que requerem confiança [10]. Geralmente, a confiança necessita de informações e baseia-se nas expectativas de um indivíduo acerca do desempenho de alguma coisa e é difícil constituir um significado base para uma noção tão ampla como confiança, uma vez que, ela pode transmutar-se de acordo com a situação ou contexto e pode basear-se em experiências e aprendizagem social, podendo, portanto, variar de acordo com nacionalidade e/ou cultura específica [10]. Porém, é preciso compreender a confiança em contextos colaborativos com os robôs, pois afeta diretamente o sucesso na utilização dos mesmos.

Como a confiança e a percepção de segurança estão diretamente interligadas, melhorar uma significa melhorar a outra. A segurança humana é um fator

importante no campo mais amplo de HRI, uma vez que facilita sua coexistência harmônica [11]. Garantir segurança em HRI não é tão simples de se resolver, antes de tudo é imprescindível deixar claro o que significa essa segurança e quais métricas pretendem ser avaliadas. Além disso, é importante intuir que cada área de aplicação da robótica permeia questões específicas a seu propósito, com normas, padrões e métricas muitas vezes diferentes entre si. Em suma, investigar e medir a confiança e as percepções humanas de segurança continuam sendo tarefas desafiadoras do domínio de Interação Humano-Robô.

Visto que os requisitos têm um papel importante a desempenhar no fator de Confiança, algumas técnicas podem servir para auxiliar a análise de restrições que contribuem para elaborar tais requisitos. Dentre elas a técnica *System-Theoretic Process Analysis* (STPA) [12] que baseia-se na teoria dos sistemas, permitindo identificar mais fatores causais e cenários de risco, principalmente os que relacionam-se ao *software, design* do sistema e comportamento humano do que as outras técnicas desse tipo. A técnica começa definindo os acidentes a serem considerados, os riscos associados a esses acidentes e as restrições de segurança do sistema. Em seguida, ações de controle inseguras em cada *loop* da estrutura de controle são identificadas e as possíveis causas para controles inseguros são determinados considerando possíveis falhas nas entradas, algoritmo de controle, modelo de processo, saídas ou *feedback* recebido pelo controlador. O STPA-Engineering for Humans [13] é uma extensão do STPA que poderá ser útil na identificação de cenários causais relacionados às interações entre humanos e automação focado principalmente no fator humano e seu comportamento.

3 Metodologia

O presente trabalho será conduzido em três etapas: levantamento dos trabalhos relacionados e taxonomias existentes, desenvolvimento do Catálogo seguido por uma avaliação dele, como mostra a Figura 1.

A taxonomia servirá como base para que a técnica de STPA possa derivar ameaças a confiança. No levantamento realizado, foi encontrado no trabalho de Langer et al. (2019) [14] diretrizes de projeto para fomentar a confiança e métodos para medi-la nas interações entre humanos e robôs (SARs) na reabilitação. Ademais, não foi encontrada nenhuma taxonomia específica de Confiança em SARs, por isso, foi preciso criar uma harmonização das taxonomias existentes em HRI (fatores que influenciam a confiança, segurança ou métricas gerais) e criar uma taxonomia adaptada mais delimitada e precisa na área a ser estudada.

Para analisar as ameaças a confiança e derivar os requisitos que podem ajudar outros pesquisadores a desenvolver e validar robôs mais confiáveis em SARs, a técnica de STPA precisa ser modificada para levar em consideração a extensão criada por France (2017) [13] e peculiaridades da Interação Humano-Robô em SARs, sendo possível, assim, derivar ameaças a Confiança mais concisas.

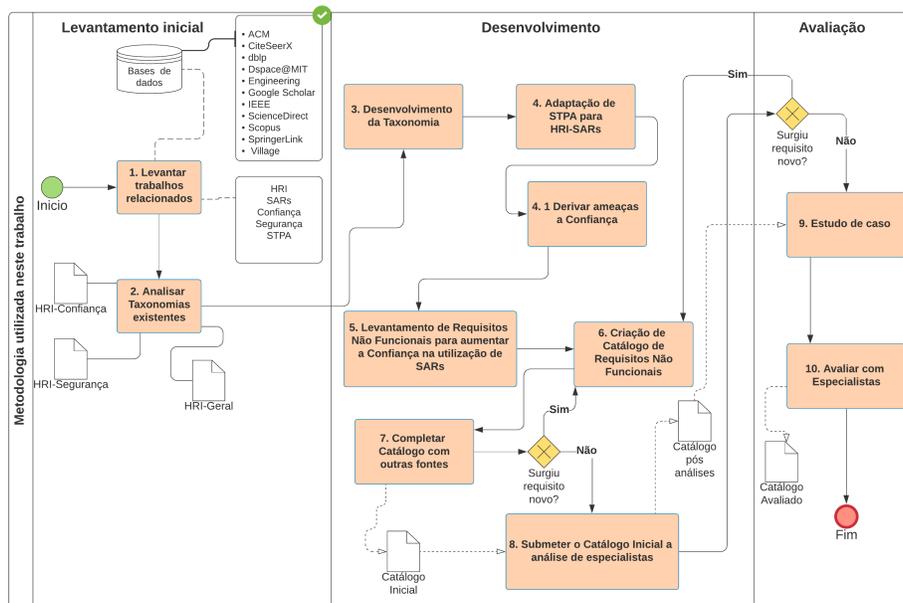


Figura 1. Metodologia utilizada neste trabalho.

4 Método de Avaliação de Resultados

Os resultados serão avaliados em três etapas essencialmente; a primeira submeterá o Catálogo inicial a análise de especialistas com perfis de conhecimento nas áreas de: Robôs Socialmente Assistivos, Confiança e Requisitos, eles responderão um questionário acerca da qualidade e completude do Catálogo. Após a análise do primeiro questionário, o Catálogo será modificado de acordo com as respostas e sugestões dadas.

A segunda etapa consiste em um estudo de caso, onde será estudado o robô NAO aplicado a fisioterapia, verificando quais requisitos do Catálogo são cumpridos, e, de acordo com os usuários envolvidos (paciente e fisioterapeuta), qual o nível de Confiança atribuído ao robô, e questionar aos usuários se no caso em que os requisitos que não são atendidos pelo Catálogo fossem implementados, eles sentiriam mais Confiança na sua utilização. Essa captação da percepção de Confiança dos usuários será realizada também através de um questionário. No fim será ratificada a viabilidade do Catálogo através de um outro questionário, com profissionais da área que não participaram da primeira etapa.

5 Estado Atual do Trabalho

A primeira fase do trabalho de aquisição de conhecimento e levantamento da taxonomia existente foi finalizada. O levantamento bibliográfico permitiu a visualização do estado da arte de HRI, SARs e STPA. Na segunda fase, a adaptação

da taxonomia com os principais conceitos encontrados na literatura foi feita e o STPA foi estruturado ao nosso propósito. Atualmente, estamos trabalhando no levantamento dos requisitos para a criação do Catálogo.

6 Contribuições Esperadas

Este trabalho tem como objetivo principal a criação de um Catálogo de Requisitos Não Funcionais que auxiliará no processo de descoberta de requisitos que incrementem a confiança humana em Robôs Socialmente Assistivos. A taxonomia gerada e a adaptação da técnica de STPA que são objetivos secundários para chegar ao objetivo principal também podem ser consideradas como contribuição para o tema.

Esperamos que o nosso Catálogo possa ser utilizado por desenvolvedores e projetistas de SARs como um guia de processo inicial de desenvolvimento, que proporcione e aumente a aceitação e sucesso da utilização dos mesmos. O NFR4TRUST poderá não só beneficiar a fase inicial, podendo também servir de modelo referência para avaliar o nível de Confiança de robôs já estabelecidos no mercado, bem como, contribuir para criação de futuras normas específicas na área.

7 Comparação com Trabalhos Relacionados

Não foi encontrado na literatura nenhum catálogo de NFR específico para SARs. Por isso, esse trabalho agregou propriedades gerais de algumas pesquisas que lidam com a criação de catálogos em outras áreas. Como no trabalho de Silva (2019) [15] onde foi criado um Catálogo de Requisitos Não Funcionais para sistemas embarcados, o de Aló (2009) [16] que traz um Catálogo de Transparência no contexto e no trabalho de Portugal et al. (2017) [17] que usam a notação do NFR de Chung que é uma representação sistemática e global de Requisitos Não Funcionais e mostram o *Softgoal Interdependency Graph* (SIG) de Transparência. Os principais trabalhos que contribuíram na criação da taxonomia adaptada para a nossa pesquisa foram Hancock et al. (2011) e Schaefer (2013) [18,19] relacionados a confiança; Murphy et al. (2013) [20] com uma taxonomia geral; Zacharaki et al. (2020) [11] relacionado a taxonomia de segurança, todos eles dentro de HRI; Já em Schaefer et al. (2016) [21] o trabalho aborda a confiança porém dentro da Interação Humano-Automação.

8 Agradecimentos

Os autores agradecem ao apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

Referências

1. Binotte, V. P.: Desenvolvimento de um Robô Socialmente Assistivo Com Controle Baseado em Comportamento de Seleção de Ação para Interação com Crianças com TEA. Programa de Pós-Graduação em Engenharia Elétrica (2018).
2. Yan, H., Ang, M.H., Poo, A.N.: A Survey on Perception Methods for Human–Robot Interaction in Social Robots. *International Journal of Social Robotics* **6**(1), 85-119 (2014).
3. Lasota, P.A., Fong, T., Shah, J.A., et al.: A survey of methods for safe human-robot interaction. *Found. Trends Robot.* **5**(4), 261–349 (2017).
4. Lee, J. D., See, K. A.: Trust in automation: Designing for appropriate reliance. *Human Factors*, **46**, 50–80 (2004).
5. Salem, M., Dautenhahn, K.: Evaluating trust and safety in HRI: Practical issues and ethical challenges. *Emerging Policy and Ethics of Human-Robot Interaction* (2015).
6. Atkinson, David John, and Micah Henry Clark.: Autonomous agents and human interpersonal trust: Can we engineer a human-machine social interface for trust?. *AAAI Spring Symposium: Trust and Autonomous Systems*(2013).
7. Cysneiros, L. M., E. Yu, and J. C. S. P. Leite.: Cataloguing non-functional requirements as softgoal networks. *Proceedings of the REAA Workshop at the 11 th Requirements Engineering Conference* (2003).
8. Feil-Seifer, David; Mataric, Maja J.: Defining socially assistive robotics. In: *9th International Conference on Rehabilitation Robotics. ICORR. IEEE.* p. 465-468 (2005).
9. Dautenhahn, K.: Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**(1480), 679-704 (2007).
10. Blomqvist, K.: The many faces of trust. *Scandinavian Journal of Management*, **13**(3), 271–286 (1997).
11. Zacharaki, A., Kostavelis, I., Gasteratos, A., Dokas, I.: Safety bounds in human robot interaction: A survey. *Safety Science* (2020).
12. Leveson, N., Thomas, J.: *STPA HANDBOOK* (2018).
13. France, Megan Elizabeth.: *Engineering for humans: A new extension to STPA.* Diss. Massachusetts Institute of Technology (2017).
14. Langer, Allison, et al.: Trust in socially assistive robots: Considerations for use in rehabilitation. *Neuroscience & Biobehavioral Reviews*, **104**, 231-239 (2019).
15. SILVA, Reinaldo Antônio da.: *Nfr4es: um catálogo de requisitos nao-funcionais para sistemas embarcados.* MS thesis. Universidade Federal de Pernambuco (2019).
16. Aló, Claudia Cappelli, and JCS do P. Leite.: *Uma abordagem para transparência em processos organizacionais utilizando aspectos.* Rio de Janeiro (2009).
17. Portugal, Roxana Lisette Quintanilla, et al.: Is There a Demand of Software Transparency?. *Proceedings of the 31st Brazilian Symposium on Software Engineering* (2017).
18. Hancock, Peter A., et al.: A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* **53**(5), 517-527 (2011).
19. Schaefer, Kristin.: *The perception and measurement of human-robot trust.* (2013).
20. Murphy, R. R., and Schreckenghost, D.: *Survey of Metrics for Human-Robot Interaction.* 197–198 (2013).
21. Schaefer, Kristin E., et al.: A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors* **53**(3), 377-400 (2016) .