Report from Dagstuhl Seminar 24292

Improving Trust between Humans and Software Robots in Robotic Process Automation

Adela del Río Ortega*[1], Andrea Marrella*[2], Hajo A. Reijers*[3], and Adriana Wilde*[4]

1 **University of Sevilla, ES.** adeladelrio@us.es
2 **Sapienza University of Rome, IT.** marrella@diag.uniroma1.it
3 **Utrecht University, NL.** h.a.reijers@uu.nl
4 **University of Southampton, GB.** a.wilde@soton.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24292 "*Improving Trust between Humans and Software Robots in Robotic Process Automation*". The seminar dealt with topics targeted at developing frameworks and guidelines to empower the trust relationship between humans and Software Robots (SW) in Robotic Process Automation (RPA).

RPA is a maturing technology that sits between the fields of Business Process Management (BPM) and Artificial Intelligence (AI). RPA allows organizations to automate high-volume and repetitive tasks – also referred to as routines – performed by human users. The enactment of these routines is emulated by means of a software (SW) robot that works on the applications' user interfaces (UIs) in the same way as the original human operators did. Recent research studies conducted on the effectiveness of RPA within organizations have found that implementation of SW robots does not always lead to the assumed effect, and many SW robots are subsequently withdrawn. In consequence, the human workforce takes over robotized tasks to perform them manually again and, in practice, replaces back SW robots. The fact is that integrating RPA into a human workforce alters the role of human employees and dynamics within the workforce, fueling a lack of trust in RPA technology, an issue deemed increasingly significant given its widespread use in many working domains.

In this direction, this Dagstuhl Seminar aimed to bring together leading experts from industry and academia engaged in diverse communities related to RPA, including BPM and Human-centered AI, intending to reflect on the current RPA principles, which fail to deliver sufficient attention to the interplay between the human workforce and SW robots. The overall goal was to explore the scientific and technological foundations to pioneer new trust-aware RPA solutions that work in partnership with the human workforce, to enhance human capabilities rather than replace human intelligence and break through the barriers to human trust using RPA. The seminar outcomes will serve as a basis to foster joint research efforts and collaborations for charting a roadmap for future RPA research.

Seminar July 14–19, 2024 – https://www.dagstuhl.de/24292
2012 ACM Subject Classification Applied computing → Business process management; Human-centered computing
Keywords and phrases business process management, human-centered AI, human-computer interaction, robotic process automation, software robots
Digital Object Identifier 10.4230/DagRep.14.7.52

* Editor / Organizer

Improving Trust between Humans and Software Robots in Robotic Process Automation, *Dagstuhl Reports*, Vol. 14, Issue 7, pp. 52–80
Editors: Adela del Río Ortega, Andrea Marrella, Hajo A. Reijers, and Adriana Wilde

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Adela del Río Ortega (University of Sevilla, ES, adeladelrio@us.es)
Andrea Marrella (Sapienza University of Rome, IT, marrella@diag.uniroma1.it)
Hajo A. Reijers (Utrecht University, NL, h.a.reijers@uu.nl)
Adriana Wilde (University of Southampton, UK, a.wilde@soton.ac.uk)

This summary provides an overview of the outcomes of our Dagstuhl Seminar "*Improving Trust between Humans and Software Robots in Robotic Process Automation*" (24292). It began with a general introduction to the aim, scope, and context of the Dagstuhl Seminar. The preliminary presentation was followed by a sequence of three invited talks required to clarify the main aspects investigated during the seminar. Specifically, Simone Agostinelli presented an "Introduction to Robotic Process Automation", and Piercosma Bisconti introduced the basic frameworks to specify trust in autonomous systems in the talk "How to build trust between intrinsic and perceived trustworthiness". Then, a third talk by Michael Rosemann discussed the existing research efforts to integrate trust into the BPM discipline in his talk: "Managing Trust in Business Processes".

The background talks were followed by three-minute speeches, during which every participant could give a brief overview of their background, expertise, and personal expectations for the seminar. The seminar participants included experts from the BPM, RPA, HCI and Trust-aware AI communities, with industry representatives and researchers in academia at different levels of seniority.

Then, a plenary brainstorming session began, moderated by the seminar organizers. Every participant in the seminar could suggest a topic they would like to investigate together with the other attendees. The discussion resulted in a set of twenty arguments that were finally distilled together into the five research questions below:

- How trust is formed between humans and RPA technology?
- How to effectively calibrate and manage trust in RPA to align pre-implementation expectations with post- implementation realities?
- How to design trust-aware human-robot interaction in RPA?
- How to identify the right balance between autonomy and trust in RPA systems?
- How to guide organizations in creating a hybrid workforce of SW robots and humans?

From the second day on, time was mostly devoted to working groups, each discussing one of the research questions that were raised in breakout sessions. This activity was accompanied by "lightning talks", i.e., brief presentations on issues that emerged during the discussions. Periodic intermediate group presentations reported on the advancement of the teams' work. The final presentations of the results achieved by the working groups and a discussion of the continuation of the freshly established collaborations took place on the final day. The seminar was concluded with the plan for writing a common vision paper to establish the foundations of a new breed of trust-aware systems.

The rest of the report contains all the talks held on the first day of the seminar, the lighting talks and the final working group reports.

2 Table of Contents

3 Overview of Talks

3.1 Introduction to Robotic Process Automation

Simone Agostinelli (Sapienza University of Rome, IT)

Robotic Process Automation (RPA) is a maturing technology in the field of Business Process Management (BPM) that creates software (SW) robots to partially or fully automate rule-based and repetitive tasks (or simply routines) performed by human users in their applications' user interfaces (UIs). Successful usage of RPA requires strong support from skilled human experts, from detecting the routines to be automated to developing the necessary executable scripts for enacting SW robots. In this talk, I explain the basics of RPA and what it means to automate a routine using a SW robot. Then, I discuss how the Robotic Process Mining (RPM) pipeline can be leveraged to minimize the manual and time-consuming steps required to create SW robots, enabling new levels of automation and support for RPA. Finally, I explore new research directions in the emerging field of AI-augmented BPM tailored to RPA.

3.2 How to build trust between intrinsic and perceived trustworthiness

Piercosma Bisconti (Interuniversity Consortium f. Computer Sci. – Rome, IT)

Trustworthiness is a complex concept and, for what concerns AI systems, its meaning and operationalization are still debated. It is unclear what does it mean to trust a non-human agent, if it is the same trust we have for humans, and how to measure it. This talk focuses on the different dimensions of trustworthiness in artificial agents, underlining the challenges for RPA to formalize what does it mean to "trust" RPA tools.

3.3 The Moral Dimension of Trust

Glenda Hannibal (Paris Lodron Universität Salzburg, AT)

Various accounts of trust have been provided over time, and the ones lead by current discussions in the philosophy of trust tend to stress the moral dimension. After providing some main distinctions in the analysis of trust, a more detailed presentation is provided of the "goodwill" trust account by Baier (1989). From this quick introduction, it becomes clear that the role of vulnerability is central to such normative requirement of interpersonal trust. In the end, the emphasis on vulnerability is used to develop a conceptual model and definition of interpersonal trust, which can be used as a starting point for empirical studies on trust in human-robot interaction.

3.4 Stop calling it autonomous if it is only automated

Christian Janiesch (TU Dortmund, DE)

Joint work of Christian Janiesch, Michael Rosemann

The talk emphasizes the critical distinction between truly autonomous systems and merely automated ones. It highlights the importance of accurate terminology to avoid misconceptions and overestimations of current technological capabilities. By clarifying these differences, the talk aims to foster better understanding and realistic expectations.

3.5 Socio-human factors in RPA: What literature says and what company does

Andrés Jiménez Ramírez (University of Sevilla, ES) and José González Enríquez (University of Sevilla, ES)

This talk integrates insights from two recent papers on Robotic Process Automation (RPA). The first study [1] explores the socio-human implications of RPA projects, highlighting the significant influence of organizational factors on project success. It presents a systematic mapping of 56 studies, revealing 16 positive and 6 negative socio-human implications of RPA, alongside 6 positive and 13 negative human inputs affecting RPA implementations. The second paper [2] discusses the evolution of RPA towards a hybrid model, emphasizing human-robot collaboration over end-to-end automation. It proposes an iterative method considering technical, psychological, and governance aspects, validated in real-world processes, showing substantial efficiency benefits. This comprehensive overview aims to provide a nuanced understanding of RPA's impact on both organizational dynamics and process efficiency.

References

1 Harmoko Harmoko, Andrés Jiménez Ramírez, José González Enríquez, and Bernhard Axmann. Identifying the Socio-Human Inputs and Implications in Robotic Process Automation (RPA): A Systematic Mapping Study. In *Business Process Management: Blockchain, Robotic Process Automation, and Central and Eastern Europe Forum*, pages 185–199, 2022.
2 Rafael Cabello Ruiz, Andrés Jiménez-Ramírez, María José Escalona Cuaresma, and José González Enríquez. Hybridizing humans and robots: An RPA horizon envisaged from the trenches. *Comput. Ind.*, 138:103615, 2022.

3.6 Trust in RPA: Between techno-empowerment and re-manualization

Artur Modlinski (University of Lodz, PL)

Robotic Process Automation (RPA) is often touted as a panacea for contemporary business challenges, promising to slash expenses, enhance product quality, and boost customer satisfaction. However, the actual outcomes in the corporate world frequently fall short of

these lofty expectations. The deployment of automated systems doesn't invariably yield the anticipated benefits, leading some firms to abandon their robotic solutions. As a result, a curious trend has emerged: human workers are reassigned to tasks previously handled by machines, effectively supplanting the automated systems. We've dubbed this reversal "re-manualization." Surprisingly, most organizations remain oblivious to this possibility until they encounter it firsthand [1]. To our knowledge, this phenomenon has yet to be thoroughly examined or documented in existing research. This widespread lack of awareness regarding re-manualization is not without consequences. It potentially exposes companies to unforeseen risks and could prove detrimental to their operations and overall performance. The gap between RPA's promised benefits and its real-world implementation warrants closer scrutiny to better prepare businesses for the complexities of automation integration

References

1 Artur Modliński, Damian Kedziora, Andrés Jiménez Ramírez, and Adela del Río-Ortega. Rolling Back to Manual Work: An Exploratory Research on Robotic Process Re-Manualization. In *Int. Conf. on Business Process Management*, pages 154–169, 2022.

3.7 Managing Trust in Business Processes

Michael Rosemann (Queensland University of Technology – Brisbane, AU)

The economy is becoming more trust-intensive as the dependence on data, the sophistication of technologies and the extent of online, purely digital experiences is increasing. As a result, we are at a tipping point where the absence of trust is becoming a significant barrier to the acceptance and adoption of technologies. This is a significant challenge as the professionalization of trust, the explicit design of trustworthiness and trust literacy, is at its infancy. This presentation offered a trust equation consisting of (1) promise, (2) uncertainty, (3) vulnerability), (4) confidence and (5) benevolence as a way to differentiate five areas of concern tomorrow's trust designers need to address when building trusted, RPA-supported business processes.

3.8 Conversational RPA and Augmented BPM

Barbara Weber (Universität St. Gallen, CH)

In my lightning talk I showed how generative AI is already used in existing automation technologies with the promise to increase automation efficiency and support democratization as users can interact with processes in a more natural way. I refer to this as conversational RPA. I contrasted conversational RPA technologies with Augmented BPM, an approach to manage processes relying on data analytics and AI to inform process improvement decisions both at design-time and run-time.

4 Working Groups

4.1 Theories for Specifying & Assessing Trust in RPA

Piercosma Bisconti (Interuniversity Consortium f. Computer Sci. – Rome, IT, piercosma.bisconti@consorzio-cini.it)
Andrea Marrella (Sapienza University of Rome, IT, marrella@diag.uniroma1.it)
Jana-Rebecca Rehse (University of Mannheim, DE, rehse@uni-mannheim.de)
Flávia Santoro (Inteli – São Paulo, BR, flavia.santoro@uniriotec.br)
Pnina Soffer (University of Haifa, IL, spnina@is.haifa.ac.il)

This working group focused on foundational issues of trust in RPA: How we can conceptualize trust between users and software robots and specify it in a way that, eventually, we can measure that trust empirically.

4.1.1 Discussed Problems

Trustworthiness in RPA is often associated with the system's technical characteristics, such as robustness and accuracy. Conversely, we believe that end-user perception of trust in RPA must be considered crucial for ensuring adoption and proper use of this technology [23]. The lack of trust arises when specific executions of an RPA system, i.e., the *SW robots*, are not entirely predictable or controllable during the enactment of specific *tasks*, fueling a degree of uncertainty about potential outcomes regarding the SW robots' behaviors [26].

Based on the above considerations, the objective of this working group has been the specification of a framework to conceptualize the notion of trust and its intrinsic and perceived characteristics. We focused on the perceived trust that a human user has concerning to a SW robot executing a specific task. In addition, leveraging our framework, we propose a novel construct to measure such a perceived trust as the *willingness of the user to give up control to the SW robot*. Finally, we discuss open issues and future works.

4.1.2 Possible Approaches

The generic trust framework [7] involves three entities: the *trustor* puts trust in the *trustee* in relation to a *trust subject*. Applied to the context of RPA, the trustor is the user of the RPA system, who is supposed to collaborate with a novel software robot, the trustee. The user needs to trust the robot to conduct a given task, the trust subject, in a reliable and valid way. Our discussions revolved around this instantiated framework, as sketched in Figure 1. We then focused on the properties of the three entities that potentially impact the level of trust among them. In the following, we briefly discuss these properties and how we suggest to specify them.

For the *user*, we see two main properties:

- Inclination to use RPA: This describes the user's general attitude towards RPA technology and hence their propensity to use it at all (independent of a concrete task or robot). It might be influenced by the personality of the user and their overall attitude towards technology. One way to further assess this inclination is by means of the Technology Acceptance Model [13].

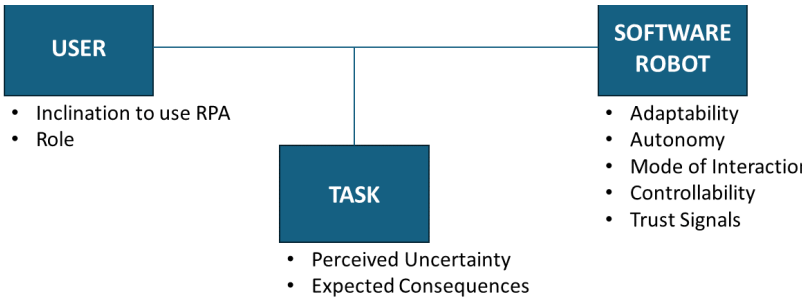

Figure 1 Trust Framework, instantiated to the RPA Context.

- Role: The other property that influences a user's trust is their role within the organization. We currently see two main roles: the (process) manager, who decides whether a software robot will be used for a certain task, and the employee, who has to collaborate with the software robot. Another possible role would be that of the developer of the bot, but that role is less relevant when considering the trust in a concrete software robot. As the developer is the one who creates that software robot, their trust needs to be placed in the underlying RPA platform, not the robot itself. Therefore, we decided not to consider it for now.

On the part of the *software robot*, we see five main properties:

- Adaptability: This describes the ability of the robot to determine its own behavior. It can be assessed by the number of task variants or behaviors that the robot can handle.
- Autonomy: We refer to Group 4 for this discussion.
- Mode of Interaction: This describes the possible ways that the user can communicate with the robot. It can be either conversational, e.g., through a chatbot, or technical, e.g., through an API.
- Controllability: This describes the ability of the user to interfere with the robot's behavior.
- Trust Signals: This describes potential trust-inducing measures that come from the environment of the robot, e.g., testimonials.

Finally, we see two main properties that are relevant for the *task*:

- Perceived Uncertainty: This describes the amount of knowledge that the user has about how the robot executes the task. In general, the more complex the task, the higher the perceived uncertainty may be.
- Expected (Negative) Consequences: This describes the negative impact that a wrong or unreliable behavior of the robot can have on the user. In general, the more important the task, the higher the expected negative consequences will be.

From these two properties, we can derive a simple classification of tasks, as outlined in Figure 2. Exemplarily, we have defined four possible classes of tasks. For each task, we have determined the degree of controllability that a user requires for this class. Both perceived uncertainty and expected consequences increase the required controllability, resulting in three different levels (low, medium, high).

To establish an indicative measure of trust between a user and a software robot for performing a given task, we make the following arguments:

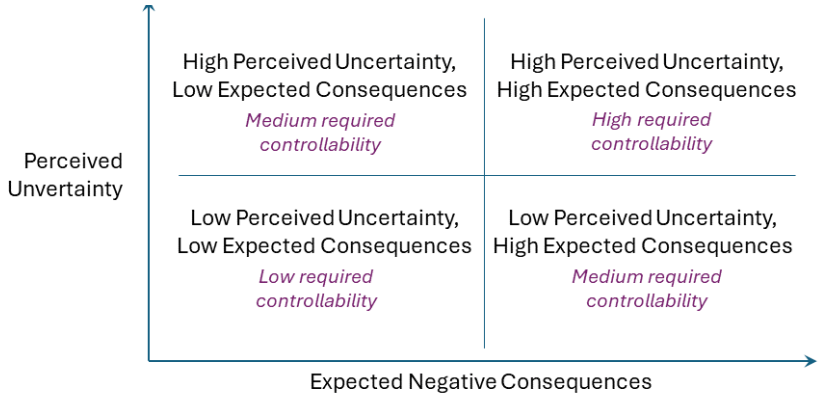

Figure 2 Possible Classification of Task for Trust Assessment.

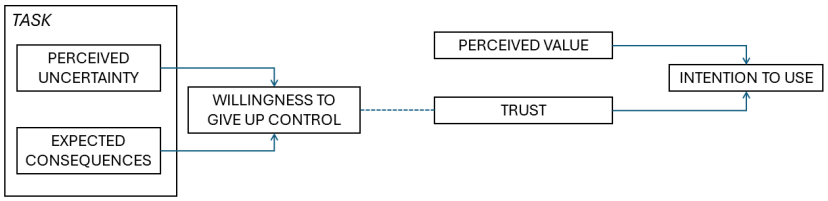

Figure 3 Trust as the Willingness to Give Up Control.

Uncertainty and possible negative consequences. The perceived risk associated with assigning a task to be executed by a software robot can be conceptualized in terms of the perceived uncertainty associated with this and the potential negative consequences that might be realized. Trust is only needed if some uncertainty (e.g., concerning the outcome, the execution status, or the time) is associated with assigning a task to a software robot. Depending on the user's role, he may still be responsible for possible negative consequences that might be incurred.

Control vs. trust. Using a software robot to execute a task does not necessarily imply that the employee trusts the software robot for this task. It is possible that control is regained by the employee by closely monitoring the software robot and by performing a thorough reviewing of the work. In this case the use does not imply a trust. When trust exists the employee gives up control over the task and trusts the software robot to execute the task, namely, the perceived risk of doing so is minimal.

It follows that the extent to which a user trusts a software robot for a given task is reflected by the extent to which the user is willing to give up control over this task. We conceptualize this as a measurable construct – *Willingness to Give Up Control (WGUC)* that can indicate the level of trust. As depicted in Figure 3, trust is independent of the perceived value of the robot. Together, they both affect the intention to use the robot.

4.1.3 Open Issues and Future Work

We propose a preliminary framework that conceptualizes how trust is formed between humans (users) and RPA (robot systems). The next step is to validate this model. Our aim is to define a comprehensive conceptual framework that precisely outlines what trust

in Robotic Process Automation (RPA) entails. This framework should encompass various dimensions of trust tailored to the interactions between users and software robots.

Future efforts should focus on creating standardized trust metrics and validation processes. These metrics need to be adaptable to different RPA contexts and capable of capturing both quantitative and qualitative aspects of trust. Identifying and validating empirical methods for measuring trust between users and software robots remains an ongoing challenge. Robust, scalable, and accurate tools and techniques are required to capture trust levels in real-world scenarios across diverse user groups.

Further exploration is needed to understand the influence of contextual factors and specific use cases on trust in RPA. Research should aim to identify how these factors impact trust and how they can be accounted for in trust assessments and frameworks. Additionally, trust is dynamic; it evolves based on user experiences and interactions over time. Understanding how trust develops, degrades, or is restored in the context of RPA is crucial. Future research should also focus on longitudinal studies to capture these dynamics and develop models that reflect the temporal aspects of trust.

Finally, as RPA increasingly integrates with artificial intelligence (AI) and other emerging technologies, future research should explore how these advancements affect trust. Understanding the interplay between RPA, AI, and user trust will be critical for developing next-generation automation systems.

4.1.4 Conclusions

By addressing the open issues and focusing on these future work areas, this research can contribute to a deeper understanding of trust in RPA and develop practical solutions to enhance trust among human users and these systems.

4.2 Trust Calibration for RPA

Simone Agostinelli (Sapienza University of Rome, IT, agostinelli@diag.uniroma1.it)
Aleksandre Asatiani (University of Gothenburg, SE, aleksandre.asatiani@ait.gu.se)
Bernhard Axmann (TH Ingolstadt, DE, bernhard.axmann@thi.de)
Glenda Hannibal (Paris Lodron Universität Salzburg, AT, glenda.hannibal@plus.ac.at)
Artur Modlinski (University of Lodz, PL, artur.modlinski@uni.lodz.pl)

4.2.1 Discussed Problems

In recent years, lightweight automation technologies such as RPA have surged in popularity. The creators of these tools have often touted them as a panacea for a myriad of business inefficiencies. With promises of unprecedented efficiency gains, cost reductions, and seamless integration across enterprise systems, RPA has fueled a wave of hype and high expectations. Promises of the eventual arrival of AI-empowered intelligent systems have only strengthened this dynamic.

However, as more and more organizations have had experience with RPA, many of them find themselves grappling with the chasm between promise and reality, as many implementations fail to deliver on the promise. This divergence often leads to mismatched expectations, resulting in disillusionment and a growing sense of **mistrust**. Therefore, it is critical to reduce the gap between the *pre-implementation* expectations and *post-implementation* reality

to ensure continued trust in the technology and eventually improve the success rate of RPA initiative.

In this working group we discussed the following research question (**RQ**): *How to effectively calibrate and manage trust in RPA to align pre-implementation expectations with post-implementation realities and ensure successful adoption and use of the technology?*

Our discussion focused on three components that we deemed necessary to address **RQ**, thus developing a successful research project:

- First, we analyzed the features and properties of RPA that contribute to its inherent trustworthiness. This included discussing attributes, such as the improved productivity, cost savings, cycle time reduction, accelerated ROI and error reduction. For a complete list of attributes please refer to the following works [2, 4].
- Then, we examined the appropriate conceptualization of the object of **trust**. This involved identifying what specifically is being trusted in the context of RPA, e.g., is the object of interest the technology or the software "agent" implemented by the RPA technology?
- Finally, we explored various methods for **trust calibration**. The discussion focused on identifying different methods for assessing the level of trust users place in the RPA. Since, existing approaches considered qualitative approaches to measure trust [31], we focused to develop a quantitative framework for trust calibration.

4.2.2 Possible Approaches

To ensure a common and strong foundation on which to further develop the trust calibration tool, we started out with a discussion about how we could use and refined the theoretical perspectives from current trust literature. This discussion lead to two main contributions:

The Gradient Trust Calibration Scale. We identified that current research on trust tends to theorize the potential trust calibration outcomes in terms of a "categorical" model [24]: trust is *warranted* (or "rational") only when either (1) people trust trustworthy persons or (2) people do not trust untrustworthy persons. The cases of misplaced trust happens in the cases where people trust untrustworthy persons or they do not trust trustworthy persons (see e.g., Table 1).

Although this categorization is helpful in understanding the aims and pitfalls of trust

Table 1 The four possible outcomes of the categorical trust calibration model.

Trust (+), Trustworthiness (+)	Trust (+), Trustworthiness (-)
Trust (-), Trustworthiness (+)	Trust (-), Trustworthiness (-)

calibration, we found this model too rigid and not directly applicable to the RPA context because we needed a theoretical model that could support a quantitative trust calibration measure. Thus, we proposed and developed our gradient trust calibration scale, which place the possible trust calibration outcomes on a continuum ranging from being very well-placed to very misplaced.

Defining the Trust Object. With this theoretical foundation, we continued by defining the trust object for the trust calibration according to the analytical framework: A trust B to C. Taking the perspective of people using the RPA, we decided to place in the analysis a human agent as A, the class of technology as B, and the specific task for the RPA to complete as C. More specifically, we discussed whether the instance of the RPA system should be a current

Table 2 Our proposed gradient trust calibration scale.

Well-placed Trust	(+)(+)
	(+)
Misplaced Trust	(-)
	(-)(-)

version without and AI extension, or it should be a future and advanced version of the technology that also utilize AI. Regardless, we took the RPA to be understood as a software "agent". From a computer science perspective, the RPA as a software agent, means that it can simply act on the behalf of someone else (e.g., human, organization). As such, our work focused on the trustworthiness of RPA as seen by people using it for task completion.

Quantification. To develop the quantitative measure of trustworthiness of RPA, we decided to develop a questionnaire. It consists of questions regarding different kinds of "trustworthiness signals" from an RPA that would be rated on the scale: -2, -1, +1, +2. Each response options in the questionnaire can then be compared to our gradient trust calibration scale to asses if a person (or group of people) fall more towards either well-placed or misplaced trust in the RPA. With the finished questionnaire, we intend to list factors constituting the perceived trustworthiness of the software robots. We suggest to construct the various aspects and decide on the selected ones by also drawing on literature related to studies of user perception of RPA. In the futher discussions, it is crucial to check how participants perceive the questionnaire and whether they have any questions or objections when it comes to its use. After the refinement of the questionnaire, we plan to validate it with our own user study by also juxtaposed it with other (more general) scales measuring trust in (general) technology.

4.2.3 Open Issues and Future Work

The discussions during the seminar provide the framing and foundation for the future research, however, there are a number open issues and action points that require more substantial work.

First, we need to conduct a comprehensive literature review on approaches to trust calibration, especially in the context of technology; and user and managerial expectations from RPA pre- and post-implementation, how these expectations are related to the inherent properties of RPA, and how these expectations evolve over time.

Second, we aim to map what are the factors that constitute the perceived trustworthiness to construct a scale measuring this construct. This step is aimed to consists of two stages. First, we want to review the literature to identify these factors. Second, we consider to make qualitative research in the home countries to add factors that somehow are missed by earlier researchers. After the factors are being constructed, we planned to make the quantitative research and to check how the scale is overlapping with other scales measuring perceived trust towards RPA.

Third, the group intends to construct a scale that would contribute to business practitioners. The analysis of the literature and business talks with business practitioners suggest that the attitudes, trust, satisfaction with RPA is not measured during annual surveys. It is a substantial shortcoming as the perceived trust influence both adoption and intention to use the system. One of the reasons why managers do not measure the trust in software

robots is the lack of a credible and easy tool that can be smoothly used in their teams. We intend to propose such a tool with the aid of both managers and actual users of the system.

To validate trust calibration scale we need to empirically test the scale and refine the scale and the eventual tool based on the empirical findings. The scale should be validated through, first the expert interviews, and then a survey. There is a potential to study trust calibration scale for two groups of human agents (as separate studies): (1) Managers making a decision about the adoption and development of RPA; (2) Users that use and collaborate with the RPA in their day-to-day work. The scale validation will require longitudinal approach surveying users at two different time points, pre-implementation and post-implementation. The trust levels will be then compared across the time points. This will allow for validating which aspects of RPA impact the trust in technology over time, and interpret the changes in attitudes of human agents over time.

4.2.4 Conclusions

In conclusion, while RPA technologies have generated significant excitement and high expectations, the reality of their implementation often falls short, leading to disappointment and mistrust. To bridge this gap, it is crucial to align expectations with actual outcomes to maintain trust and improve the success rate of RPA initiatives. Our working group focused on calibrating trust, understanding trustworthiness, and developing a trust measurement scale. By conducting comprehensive research and creating practical tools for business practitioners, we aim to ensure more reliable and sustainable adoption of RPA technologies.

4.3 Human-Robot Interaction in Relationship to Trust

Marco Angelini (Link University – Rome, IT, m.angelini@unilink.it)
Angelo Casciani (Sapienza University of Rome, IT, casciani@diag.uniroma1.it)
Christian Janiesch (TU Dortmund University, DE, christian.janiesch@tu-dortmund.de)
Antonio Martínez-Rojas (University of Seville, ES, amrojas@us.es)
Manuel Resinas (University of Seville, ES, resinas@us.es)
Adriana Wilde (University of Southampton, UK, a.wilde@soton.ac.uk)

4.3.1 Discussed Problems

One critical aspect that often undermines the efficiency and effectiveness of the interactions between humans and software robots ("robot", for short) is the calibration of trust of human users [32].

That is, when humans work with a (new) robot, their level of trust in that robot is not always appropriately calibrated. This miscalibration manifests in two primary ways: overtrust and undertrust [19]. Overtrust occurs when users place too much confidence in a robot, potentially overlooking its limitations and risks. Undertrust, conversely, happens when users are overly skeptical of a robot's capabilities, leading them to underutilize it or excessively monitor its operations, leading to degrading performance.

Calibrating trust to an appropriate level is not a straightforward process. It involves a complex interplay of various contextual factors, including user previous experiences, robot

performance, and transparency of robot operations. Trust calibration thus requires time and consistent interaction between the user and the robot.

To address the issue of miscalibrated trust, we propose varying the way human and robot interact rather than calibrating the trust of a user ex-ante. This approach involves providing users with appropriate control and observability into the interaction with the robot accepting some inefficiency rather than risking disuse or misuse.

However, in current RPA products, the automation spectrum is typically only viewed as a binary option: attended versus unattended [3]. This simplification limits effectively managing human-robot interactions and establishing warranted trust. Furthermore, it has been shown that not all attended mechanisms are the same and just asking for users validation is not enough to increase trust in the robot and, ultimately, user acceptance [30].

Therefore, the central research question driving this exploration is: *How do we design trust-aware human-robot interaction in RPA?* By answering this question, we aim to create a workable solution that can eventually enable users to develop the appropriate level of trust in the robot.

4.3.2 Possible Approaches

To address the discussed problems, we propose the following possible approaches:

Continuum of interaction. To address the limitations of the widely accepted view of RPA, where automated tasks are on a dichotomy of attended or unattended tasks, we propose a continuum ranging from fully attended to fully unattended tasks. There exists a prior art in the classification of levels of automation [17, 27]. Yet, they assume delegation task to an automated entity. With the evolution of RPA into intelligent robots, other more advanced forms of interaction become possible and have not been considered so far. The continuum reflects the levels of human involvement required in a given work system, as depicted in Fig. 4. The degree of involvement from both the human and the robot (or "machine") determines the following levels of interaction:

- **Machine out of the loop** (*MootL*), in other words, there is no robot involvement, other than *ex-post* reviews, as the tasks are completed in a fully manual manner.
- **Machine on the loop** (*MotL*): The human is still the performing agent, but there might be some *ex-durante* robot oversight such as in the case of a machine instructor.
- **Machine in the loop** (*MitL*): The human performs the task, but in addition to oversight from a robot, there is also some *ex-ante* control by the robot, such as in the form of an authorisation. This, and the previous level are hypothetical and included here for completeness of the model but not typically observed in the wild in the context of RPA (hence grayed out in Fig. 4), as any move from hybrid intelligence towards remanualization is not typically done through these levels, but a rather abrupt leap (seen as the curved arrow in Fig. 4.)
- **Hybrid intelligence:** This is the "inflection point" reflecting increasing automation in a manner where the human is no longer the sole performative agent with regards to the task, but robots assume some of these roles.
- **Human in the loop** (*HitL*): In this case, the robot performs the tasks with *ex-ante* interventions, as in the case of those tasks requiring human authorization before the robot can proceed.
- **Human on the loop** (*HotL*): The robot has an increased level of automation. Yet, for tasks yet not fully unattended there is the possibility for *ex-durante* human intervention

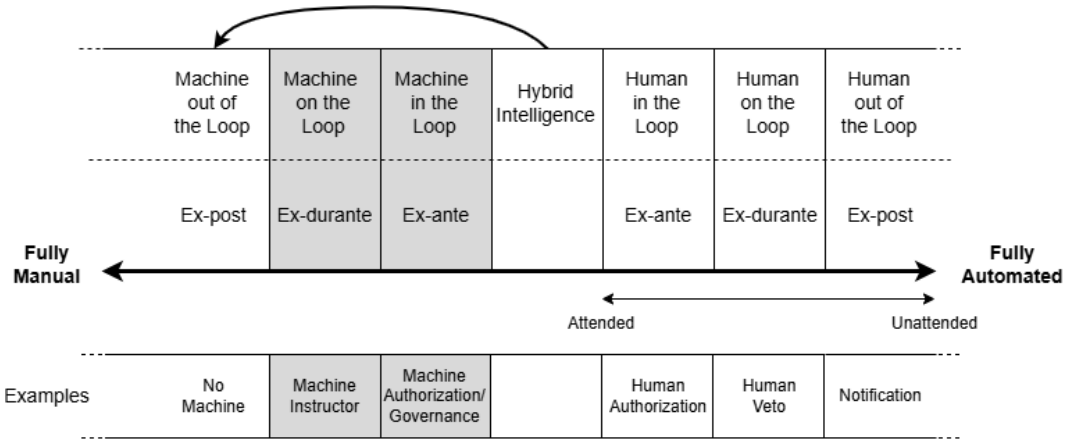

Figure 4 Human-robot interaction levels across the automation spectrum.

whilst the robot performs the task, for example through vetoing.

- **Human out of the loop** (*HootL*): This is the case of a fully automated task that is unattended, but the user may receive *ex-durante* or *ex-post* notifications (at the robots discretion), but cannot directly interact with the robot.

Automation of operations vs. automation of supervision. So far, the distribution of work in levels of automation has focused on transferring classes of operational activities to a robot while keeping the human user either in, on, or out of the loop as a process participant and/or supervisor at the same time. The role of the process participant and supervisor could be distributed to two humans. Conversely, the robot could assume the supervisory role and be the machine in or on the loop in a supervisory role. That is, the robot could require the human to request *ex-ante* authorizations and itself exert *ex-durante* interventions in human and/or robot operations. We have included this trail of thought in Fig. 4 by including levels of MitL and MotL shaded in gray.

Classes of activities. To implement these adaptive models, we re-use four classes of activities that form the interaction design space in Fig. 5 [27]:

- Information acquisition/sensing: This class involves gathering necessary data and resources to understand the task or problem from reliable sources. Effective acquisition ensures the accuracy and completeness of data, setting a solid foundation for subsequent analysis and decision-making processes.
- Analysis: The analysis class involves interpreting gathered information to identify patterns and generate potential solutions. The data is broken down and viable action candidates are being developed. This class bridges information acquisition and decision-making, transforming raw data into meaningful insights for informed decision-making.
- Decision: This class involves selecting the best course of action from identified action candidates. It considers analysis results, potential impact, feasibility, and risks. The goal is to choose the action that maximizes benefits while minimizing drawbacks.
- Action: This class implements the selected decision through actions. Success can be measured by achieving desired objectives and solving the initial problem.

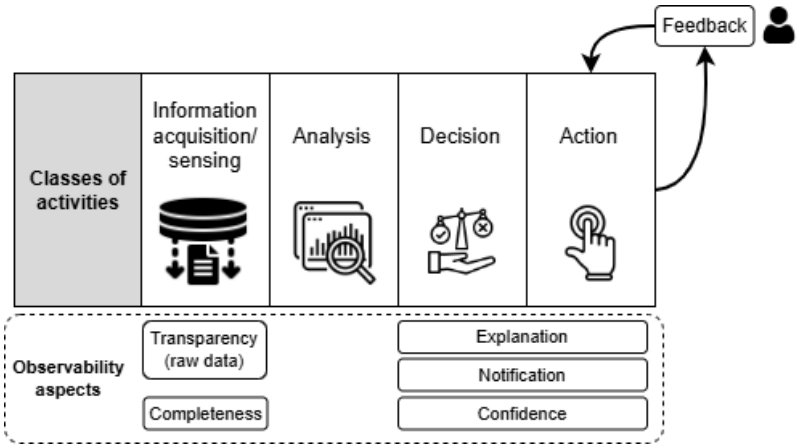

Figure 5 Classes of activities and observability aspects.

Increasing observability. A key meachanism to improve the trustworthiness of a robot is increasing the observability of the activities it performs. For instance, in [30], the authors show how they designed their robots to include explanations about the process they followed, the uncertainty about the work they performed, and the raw data they processed to increase the trust of the users in its behavior.

There are different aspects that can be observed in the behavior of a robot that can be used to increase its trustworthiness. Taking inspiration from some of the principles of trustworthy AI defined in [20] together with the use cases described in [30], we provide an initial classification of these aspects, which is also depicted in the lower part of Fig. 5:

- **Explanation:** It includes providing explanations about how the robot has made the decision or how it has performed an activity. For example, a robot that searches for a document in several information systems could provide information about the repositories it searched for and where the document was found.
- **Notification:** It includes notifying the user about the result of the action, maybe including the errors that have occurred during the process. It may also refer to notifying the user about the decisions made in the process. Finally, it also includes providing aggregated information about the percentage of cases in which the robot worked properly.
- **Confidence:** It includes providing the user with an estimation of the level of confidence of the robot about the decision or the action performed. This makes sense particularly in those cases where the activity performed requires a certain degree of intelligence. For instance, if the robot needs to perform an OCR task on an invoice to extract some information from it, the robot could provide a numerical score that captures how confident it is on the action performed.
- **Transparency:** It includes providing the user with the raw information the robot used to perform the task. For instance, in the previous example of the OCR task, this would mean giving the user access to the scanned invoice that was used by the robot.
- **Completeness:** It includes providing the user with information about whether it could fulfill the task completely or if it could not fully complete it and could require some additional information to do the task.

We believe that the choice of which observability aspects to consider in the design of the robot and whether the robot should provide the information reactively or proactively depends heavily on the task that the robot is performing, as well as other elements of the context such as previous experiences with RPA or organizational culture. Finally, in a classical RPA setting, the list of observability aspects we have collected involves an information

flow from robot to human. However, we believe that it is also convenient to consider the information flow in the opposite direction. In this sense, we have identified *feedback* (see the upper part of Fig. 5) as a relevant aspect. The feedback provided by users may give useful information to the robot to improve its behavior. This improvement could be done automatically. For example, an intelligent robot could use this feedback to retrain its machine learning model and improve the accuracy of its predictions. It could also be useful in a more manual setting, in which the developer of the robot uses this information to improve or extend its behavior.

4.3.3 Open Issues and Future Work

In this section, we present open issues arising from the proposed approaches that we foresee should be coped with toward the full realization of our proposal:

Understanding the evolution of trust. Understanding how trust in the robot evolves and impacts interaction levels is crucial. This includes identifying the appropriate interaction level for different trust contexts. The context may encompass multiple factors like the task at hand, its associated risk level, or the intended persona (e.g., developer, domain expert) influencing the user's trust in the robot's task performance. It also includes defining how the degree of automation of operations (e.g., increasing/decreasing the level of automatic execution of activity) and observability (e.g., getting explanations or notifications for the human who is supervising the process) affect the trust level of the human and robot with each other.

Balancing utility and trust at the interaction level. Determining how to balance perceived utility against trust is an ongoing challenge. Our current work remains in a high-level design. Operationalizing these interactions in domain-specific ways is necessary. Designing human-robot interaction patterns could prove effective in identifying the trade-off between increased trust and lowered utility or vice-versa. Developing a pattern library of human-robot interactions and their expected effect on trust is a promising research direction.

Designing interactions for the hybrid intelligence level. Looking at human-robot interaction levels, the hybrid intelligence level is the most innovative and challenging one to design solutions for. In this level the interaction design must consider both humans and robots being capable of executing operations and supervising the process for their needed part. This scenario shows a complete collaboration between the human(s) and robot(s) with the highest frequency of interactions. Coordinating this interplay while keeping mutual trust can prove challenging while holding significant potential.

Developing a methodology for evaluating interaction design pattern efficacy. Designing a methodology to evaluate the effectiveness of interactions in addressing trust issues is another open aspect requiring further research. This challenge requires first the capability to measure in quantitative or qualitative form the relation between interaction effectiveness and resulting trust levels. Methodologies should be developed to inform the developer during the interaction design process on how to estimate the potential trust gain, how to relate it to interaction design patterns, and how to improve their design. Finally, similar considerations should be applied to assess the efficacy of the designed interactions with real users during realistic utilization scenarios.

4.3.4 Conclusions

The consideration and design of human-robot interaction can be a crucial factor in achieving correct and effective trust calibration, particularly in the realm of RPA. By providing users with appropriate control and observability in working with robots, we have devised a proposal to deal with the issue of miscalibrated trust ex-ante. We have explored this issue using the established lens of automation levels based on classes of activities. In this way, a match between the perceived trust of the user and warranted trust can be established when the user is not yet properly calibrated to the robot to improve acceptance. By accepting a reasonable amount of miscalibrated inefficiency, we envision that user can better calibrate their trust towards the system and prevent misuse and disuse of under- or overtrusted robots.

Although our focus has primarily been on human trust in robots, the discussed concepts are also applicable to exploring the reverse – robot trust in humans. This perspective is valuable when designing machine-in/on-the-loop systems, where robots monitor and support human tasks as co-pilots. This bi-directional trust paradigm could enhance future interaction design, promoting even more effective and trustworthy human-robot collaborations.

4.4 RPA Technology and Trust

Christian Czarnecki (FH Aachen University of Applied Sciences, DE, czarnecki@fh-aachen.de)
Andrea Delgado (Universidad de la República, UY, adelgado@fing.edu.uy)
José González Enríquez (University of Seville, ES, jgenriquez@us.es)
Faizan Ahmed Khan (University of Padova, IT, faizanahmed.khan@studenti.unipd.it)
Barbara Weber (University of St. Gallen, CH, barbara.weber@unisg.ch)

4.4.1 Discussed Problems

This working group focused on the development and evolution of Robotic Process Automation (RPA) and trust in these systems as they become increasingly sophisticated. We began by examining the different types of RPA technology and their characteristics. Additionally, we identified use cases and categorized them within our classification framework. We then explored whether different types of RPA technology have varying trust requirements. Instead of concentrating on specific instantiation of RPA systems, tasks, and users, we focused on different types of RPA systems and their underlying characteristics.

4.4.2 Possible Approaches

4.4.2.1 Types of RPA Technology

Table 3 shows different types of RPA technology characterized by the nature of their input data (structured versus unstructured), their decision-making (rule-based versus data-driven based on AI), and their adaptability (reprogramming versus learned). This provides the basis for framing the analysis of each type of RPA technology, including its main characteristics and trust requirements, as we move towards more sophisticated RPA technology.

Table 3 From Traditional RPA to Autonomous RPA.

	Traditional RPA	**Cognitive RPA**	**Intelligent RPA**	**Autonomous RPA**
Data	structured	unstructured	unstructured	unstructured
Decision making	rule-based	rule-based	data-driven based on ML	data-driven based on ML
Adaptability	through reprogramming	through reprogramming	through learning	through learning, continuous feedback loops, and real-time data processing
Example	Use Case 1	Use Case 2 + 3	Use Case 4	

As shown in Table 3, *Traditional* RPA is limited to structured data and predefined rules. In turn, *Cognitive RPA* is still relying on predefined rules, but is able to extract structured data from unstructured data (e.g., extract structured data from a PDF via OCR) which is passed as input to RPA bots, which then perform some actions (e.g., data entry in an enterprise system). Both traditional and cognitive RPA require reprogramming to adapt to changes (e.g., a regulatory change that requires changes to a task/process [26]). *Intelligent RPA* does not solely rely on rule-based decisions anymore but is able to conduct data-driven decisions based on ML. Adaptation does not require (re)programming but the mechanisms behind adaptation is learning. Finally, *Autonomous RPA* typically continuously senses its environment (i.e., have a continuous feedback loop), processes this data and uses real-time data to adjust.

Especially traditional RPA heavily relies on stable and consistent input data which is provided in a structured manner. If the input data changes, different challenges may arise. This can include bot failures (e.g., due to changes in data format or unexpected values), potentially even leading to process disruption, accuracy issues (e.g., inconsistent or low-quality data), maintenance overhead (e.g., changes of the bot to accommodate changes in the data including re-testing), and lead to integration challenges (e.g., with the target system). Moreover, if data sensitivity or input data changes, the bots must be updated to avoid security or compliance issues. Cognitive RPA, which relies on AI/ML for handling unstructured data, relaxed this reliance on structured data. This tends to be more robust to variations in the input as it allows to interpret and adapt to different formats (cf. Figure 6).

Orthogonal to the classification above is Conversational RPA. Conversational RPA enables users to interact with the process through text or speech (i.e., conversational interfaces), which might affect trust requirements. Moreover, any of the classes of RPA technology defined above could integrate APIs (even ML APIs and LLMs) which can increase trust requirements as the expected outcome might become harder to predict. Another relevant distinction is involvement of humans in a concrete implementation. While attended robots still require human control and intervention, unattended robots work without any human involvement. Human attendance can help to resolve potential uncertainties manually (e.g., adding missing data manually or validating data entries where the confidence of the bot is limited), mitigating certain trust issues.

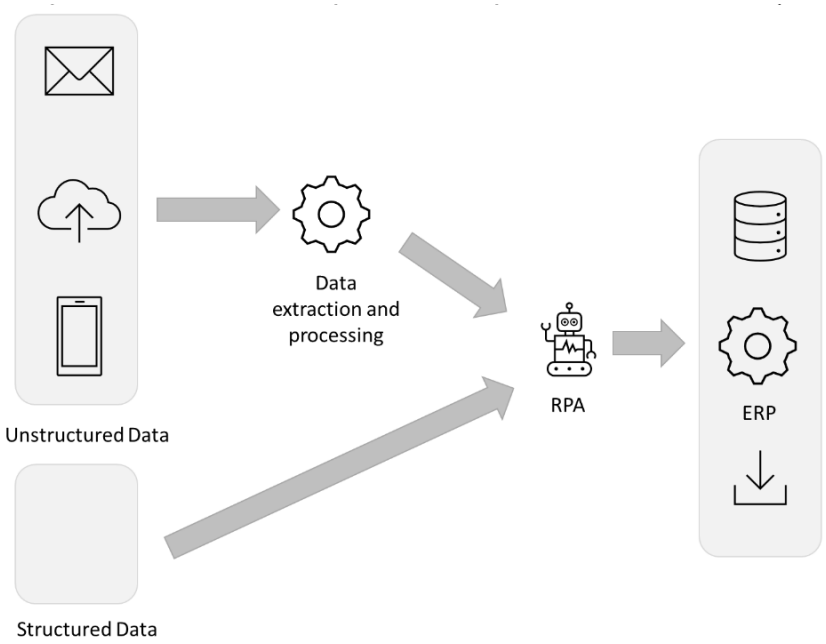

Figure 6 Differentiation between structured and unstructured input data.

4.4.2.2 Selected Use Cases as RPA Instantiations

In the following we introduce different use cases which describe concrete instantiations of RPA usage.

Use Case 1: Travel authorization request form automation (Traditional RPA). Use Case 1 refers to a traditional RPA application and automates the travel authorization process for a university (cf. [1]). The data is entered in an excel spreadsheet by the person requesting the travel authorization (structured data). The form automation records the file location (path) and the form to be filled, and the relation between each row with each field to look for the specific values for each request. The RPA bot does not involve any decision making, but just maps the data in the excel file with the user interface (see Figure 7). However, rule-based decisions could have been in place with a different implementation. Potential breaking issues in this type of automation are related to hardcoded references such as file path, field correspondence, missing values, etc. Adaptations to resolve these issues require updating the robot (adaptation through reprogramming).

Use Case 2: Electricity Consumption Data Extraction (Cognitive RPA). Use Case 2 refers to cognitive RPA application for electricity data extraction [9]. As illustrated in Figure 8, this use case has automated the manual process of extracting and booking electricity consumption data. The unstructured input data of paper invoices was transferred to structured data by using a software robot (cf. Figure 8 step 3 of process row). The cognitive aspect of this software robot was recognizing the correct data in the invoice (i.e., line, data field) and linking it to the proper entry in the ERP system. After the automated recognition, there was still a manual verification step (i.e., attended RPA). This manual verification helped to decrease uncertainty and, therefore, likely helped the user to trust the implementation. Furthermore, the accuracy before and after manual intervention was measured and reported. In the end, 100% of lines and 99% of data fields were recognized correctly. However, 52%

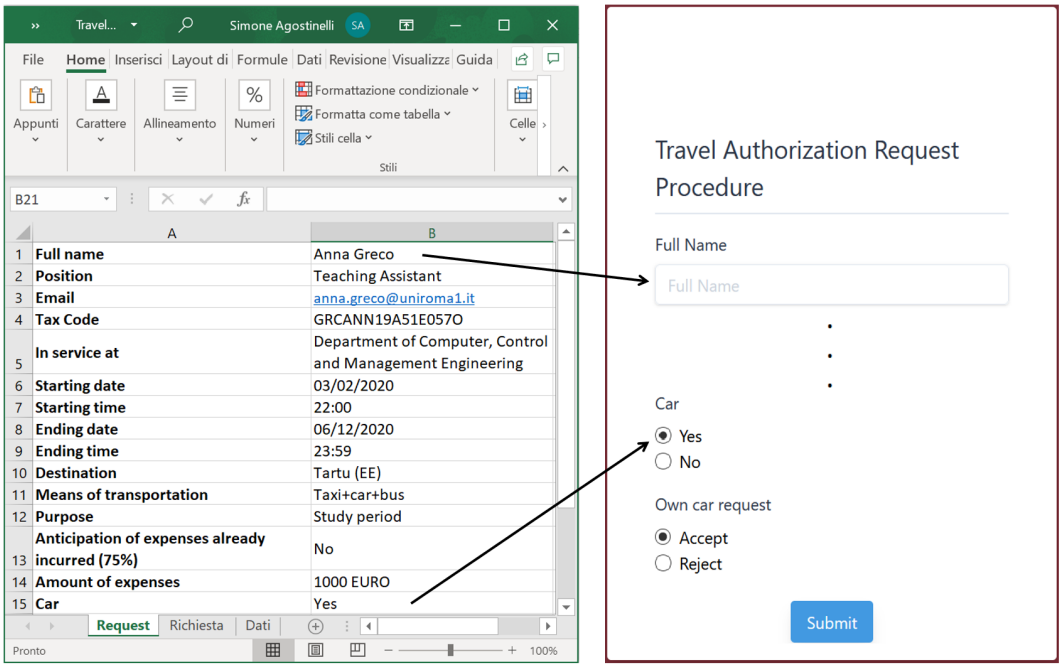

Figure 7 Example of traditional RPA for a university travel authorization request form automation [1].

of lines and 11% of data fields required manual intervention. Moreover, conversations with stakeholders showed decreased failure rates compared to the manual process. Overall, the accuracy of the software robot attended by a human was higher compared to the purely manual process. The broad accuracy of the implementation, combined with the fact that the RPA implementation outperformed the manual process, likely contributed to increased trust in the implementation.

Use Case 3: Invoice management (Cognitive RPA). Use Case 3 is another application of Cognitive RPA automating a task comprising several activities related to the management of invoices (cf. Figure 9) [30]. The process involves (1) locating the client's invoice using the invoice number across various systems (e.g., ERPs, CRMs, folders, or email) and (2) validating the data recovered from the invoice, e.g., VAT number, the amount of the invoice matching SAP records, and considering different company layouts, for instance. (3) If errors or missing information are found, the client is contacted for corrections, requiring them to restart the process. (4) If correct, the data is entered into the SAP system to confirm validation and continue the process. This use case has to deal with unstructured data (e.g., searching for invoices in different systems and processing invoices in potentially different formats). To mitigate the risk of low-quality input data retrieved by the robot, the human is kept in the loop to validate the data in case of missing or incorrect information (cf. Figure 9 activity 2) and to decide how to proceed (cf. Figure 9 activity 3 or 4). Even though this decision could be automated, it was intentionally kept manual to allow the human to feel in control over the process, which increased users' trust in the robot [30].

While implemented as a cognitive RPA application, this use case could easily evolve into an intelligent RPA application by delegating data validation to an ML system. This would presumably increase the trust requirements as the expected outcome might become harder to predict.

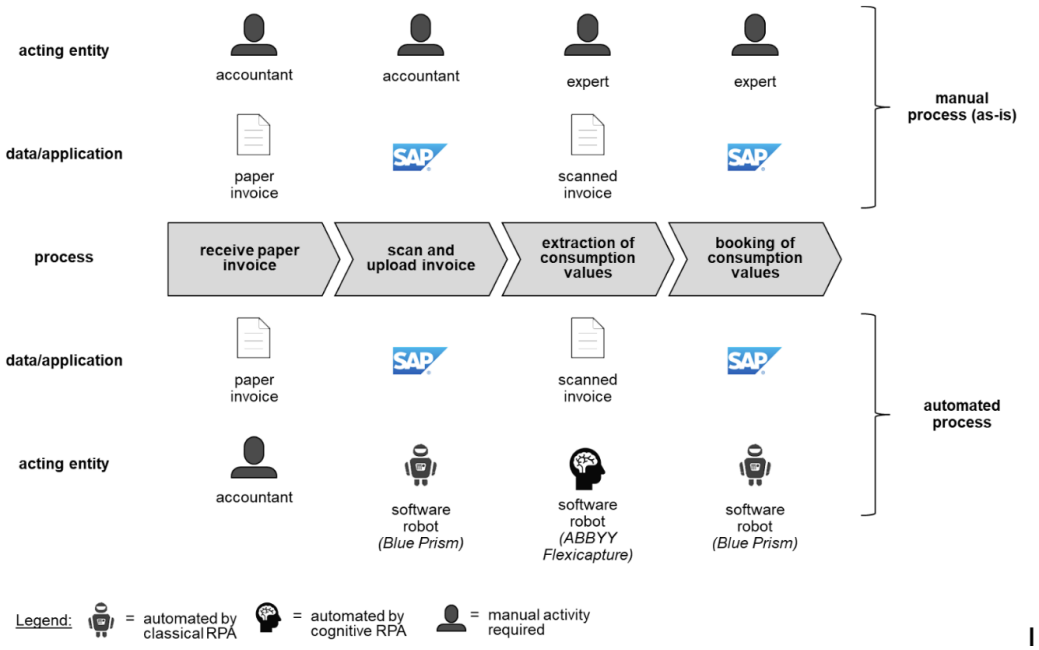

Figure 8 Example of Cognitive RPA for energy consumption data extraction [9].

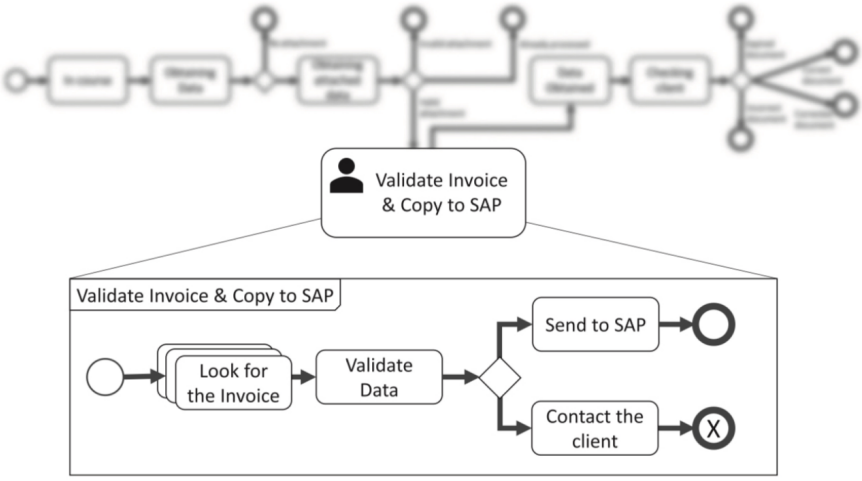

Figure 9 Example of Cognitive RPA for looking for invoices [30].

Use Case 4: Complaint handling at an electricity company (Intelligent RPA). Use Case 4 relates to the context of Business Process Outsourcing (BPO), where the process involves handling a set of claims from an electric company (see [21]). In the first activity (cf. Figure 10), a robot collects unstructured data from various applications or systems (e.g., meter readings, consumption, payment punctuality, recent bills, or customer age). Once collected, the data is processed, and a decision proposal (i.e., accept or reject the claim) is generated with an associated score determined by an ML system. Finally, all decisions

must be validated by a human operator, allowing feedback based on the human decision to be collected and used to retrain the ML system. The undeterministic nature of ML, further increases trust requirements.

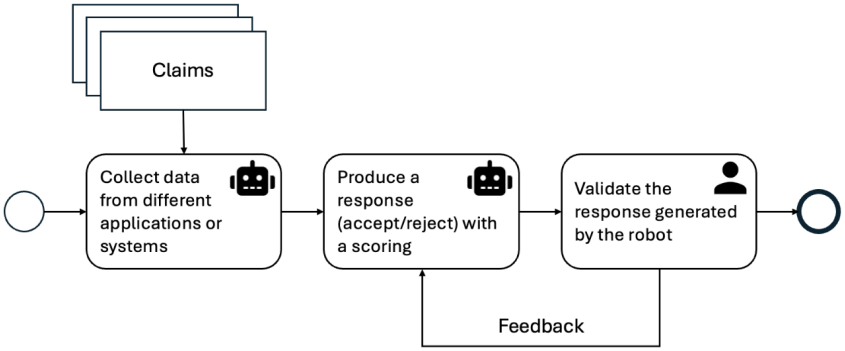

Figure 10 Example of Intelligent RPA for attending complaints of an electricity company (inspired by an industrial experience [21]).

4.4.2.3 Relation between RPA Technology and Trust

As seen in the different use cases, it can be assumed that the type of RPA technology and trust requirements are interrelated [15]. In this context, trust can be understood as "a complex addition of intrinsic and interactional characteristics" [5].

Comparing the different types of RPA (i.e., traditional, cognitive, intelligent, autonomous), we can see an increasing level of sophistication in RPA technology and increasing uncertainty. Assuming an inverse relationship between uncertainty and trust, the increasing uncertainty leads to increased trust requirements (cf. Figure 11).

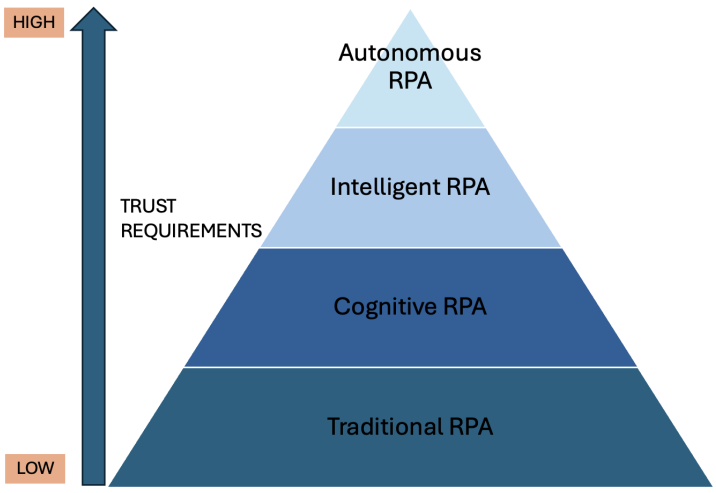

Figure 11 Trust requirements over the RPA technology classification.

To meet the trust requirements related to the different RPA types, the trustworthiness, i.e., "the ability to meet stakeholder expectations in a verifiable way" [16], needs to be

addressed. Trustworthiness characteristics can be further specified by using the AI Trustworthiness Framework [8] depicted in Figure 12, for example, accuracy, robustness, and transparency. However, the concrete representation of those requirements is mostly dependent on the automated task and its context, therefore, related to the instantiation. For instance, the AI Act [28] already defines levels of risks regarding different domains and tasks, e.g., a bot for processing student admissions or a bot for credit scoring would be considered high-stakes AI, which would impact the trust requirements for the specific instantiation.

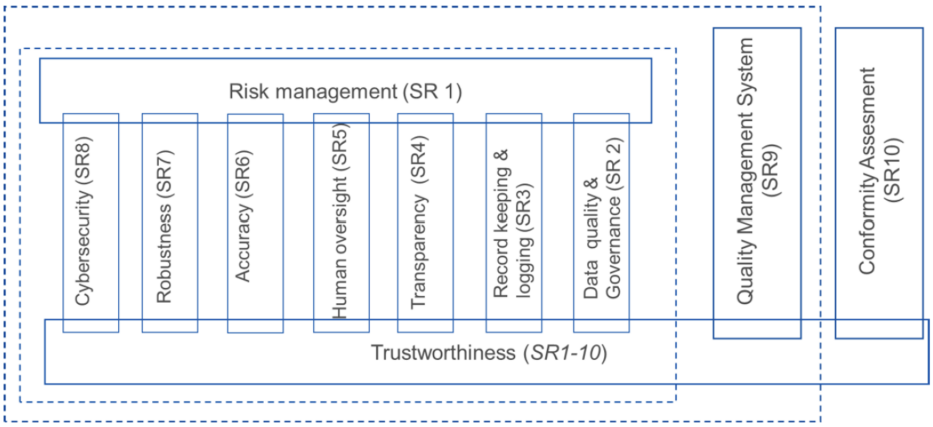

Figure 12 Architecture of standards of the AI Trustworthiness Framework [18].

4.4.3 Open Issues and Future Work

We have mainly focused on intrinsic characteristics at the RPA technology level. Several relevant elements of the relationship between trust and RPA technology are instantiation-dependent, which we discussed only for selected use cases. These elements would need to be addressed more systematically to provide a complete vision of trustworthiness. As mentioned before, this can be done using the AI Trustworthiness Framework, also considering the environment and context of the specific instantiation.

4.4.4 Conclusions

Regarding the initial question of how to create trust for RPA systems, we found an interrelation between the type of RPA technology and its trust requirements. To address this, we first structured the different technology types using the dimensions of data, decision-making, and adaptability. Consequently, we propose the following four types of RPA technology: (1) Traditional RPA, (2) Cognitive RPA, (3) Intelligent RPA, and (4) Autonomous RPA. As the sophistication of RPA increases, so does the associated uncertainty. Assuming an inverse relationship between uncertainty and trust, this rising uncertainty leads to heightened trust requirements. Specific trustworthiness characteristics can address these increased trust requirements for concrete implementations. In summary, we contribute to the overall topic of trust and RPA by providing initial insights into the intrinsic trust requirements associated with the various classes of RPA technologies.

4.5 Guiding Organizations in Creating a Hybrid Workforce of Software Robots and Humans

Adela del Río Ortega (University of Sevilla, ES, adeladelrio@us.es)
Andrés Jiménez Ramírez (University of Sevilla, ES, ajramirez@us.es)
Ralf Plattfaut (Universität Duisburg-Essen, DE, Ralf.Plattfaut@icb.uni-due.de)
Hajo A. Reijers (Utrecht University, NL, h.a.reijers@uu.nl)
Michael Rosemann (Queensland University of Technology (QUT) – Brisbane, AU, m.rosemann@qut.edu.au)

The final working group of the seminar was titled "Guiding organizations in creating a hybrid workforce of software robots and humans" and dealt with trust within a hybrid workforce containing human and robotic agents within organizations.

4.5.1 Discussed Problems

There is great agreement, that the technological capabilities of process automation solutions increase over time. Recent examples include the integration of cognitive automation [11], machine learning [25], and artificial intelligence [14]. As such, the process automation solutions become increasingly sophisticated.

Companies integrate these increasingly intelligent process automation solutions into their business processes and services [14]. Predominant goals of this automation are to improve costs, quality, time, and flexibility of the processes and services [12].

With the increased sophistication of the automation solutions integrated in processes and services, perceived ability, integrity, and benevolence changes. Anecdotal evidence includes fear of wrong decisions through AI, biases within the decisions, or malicious intents of more general AI [12]. These are, e.g., exemplified within media reports [1] [2].

As ability, integrity, and benevolence form the basis of trust [22], processes and services require a higher degree of trust than before. These trust needs can be analyzed from three different perspectives:

- Managers and decision makers require trust in the process automation solution to purchase the corresponding solution and allow organizational adoption in the first place.
- Employees within the organization need trust in the process automation solution to integrate the solution into the processes and services they are responsible for.
- Customers of the organization need trust into the automated process/service so that they confidently consume this service.

However, situations emerge, where the current trust level is not enough for the increasingly sophisticated automation solution. The simplified assumption is that increased sophistication increases uncertainties (e.g., inability to explain the system's working) and vulnerabilities (e.g., potential loss of data that now has to be provided). Here, a trust gap emerges between the current level of trust and the level of trust required by the systems.

[1] https://www.ft.com/content/979fe974-2902-4d78-8243-a0cff68e630a
[2] https://www.dw.com/en/ai-germany-discrimination/a-66670854

However, without dedicated trust building there will be a threshold at the level of the actual trust. In these situations, trust signals are needed to persuade managers, employees, or customers to still adopt, implement, or consume the technology within the service. These trust signals can be interpreted as arguments to alleviate trust. Examples include indirect experiences from peers, the opinions from experts in the field, reputation of the provider of the automation solutions [29].

4.5.2 Possible Approaches

We propose a framework for associating trust signals with trust concerns. Trust concerns are manager, employee and customer specific. Here, we start by focusing on the trust concerns of the employee by drawing a Maslow-like triangle[3] (see Fig. 13) of employee concerns regarding the introduction of RPA as part of their employer's organization, including major concerns identified by the OECD such as job security and quality[4]. These trust concerns can be addressed with specific trust signals.

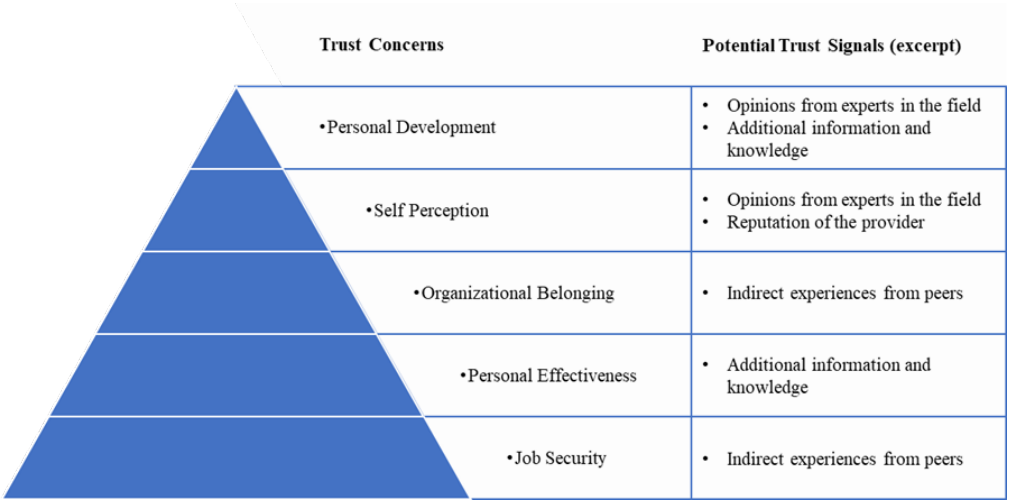

Figure 13 Trust Framework, instantiated to the RPA Context.

4.5.3 Open Issues and Future Work

Further research is required on the emergence of the trust gap and its effects within the organization. Similarly, further research is needed on the effects of different trust signals. Trust signals have been studied in prior research, e.g., in e-commerce contexts, where they were shown to change consumer behavior [10]. However, a transfer to trust within advanced process automation solutions is yet needed. Employees, managers, and consumers need to have a deeper understanding of the automated processes and services that support the building of trust [6]. The role of trust signals is to help them build this understanding.

[3] Maslow's hierarchy of needs is a motivational theory in psychology comprising a five-tier model of human needs, often depicted as hierarchical levels within a pyramid.
[4] https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2023_08785bba-en

References

1 Simone Agostinelli, Marco Lupia, Andrea Marrella, and Massimo Mecella. Reactive synthesis of software robots in RPA from user interface logs. *Comput. Ind.*, 142:103721, 2022.

2 Aleksandre Asatiani, Esko Penttinen, Joona Ruissalo, and Antti Salovaara. Knowledge workers' reactions to a planned introduction of robotic process automation—empirical evidence from an accounting firm. *Information systems outsourcing: The era of digital transformation*, pages 413–452, 2020.

3 Bernhard Axmann and Harmoko Harmoko. Robotic Process Automation: An Overview and Comparison to Other Technology in Industry 4.0. In *10th Int. Conf. on Advanced Computer Information Technologies, ACIT 2020*, pages 559–562. IEEE, 2020.

4 Bernhard Axmann and Harmoko Harmoko. The assessment model of robotic process automation (rpa) project using benefit study and balanced scorecard (bsc) approach. *Tehnički glasnik*, 18(2):246–253, 2024.

5 Piercosma Bisconti, Letizia Aquilino, Antonella Marchetti, and Daniele Nardi. A Formal Account of AI Trustworthiness: Connecting Intrinsic and Perceived Trustworthiness. In *Aies '24: Proc. of the 2024 Aaai/Acmconference on Ai, Ethics, and Society*. forthcoming.

6 C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: mental components and sources of trust. In *33rd Annual Hawaii Int. Conf. on System Sciences*, pages 10 pp. vol.1–, 2000.

7 Jin-Hee Cho, Kevin Chan, and Sibel Adali. A Survey on Trust Modeling. *ACM Computing Surveys*, 48(2), 2015.

8 European Commission. Draft standardisation request to the european standardisation organisations in support of safe and trustworthy artificial intelligence, 2022.

9 Christian Czarnecki, Chin-Gi Hong, Manfred Schmitz, and Christian Dietze. Enabling digital transformation through cognitive robotic process automation at deutsche telekom services europe. In *Digitalization Cases Vol. 2 : Mastering digital transformation for global business*. 2022.

10 David Dann, Raphael Müller, Ann-Catherin Werner, Timm Teubner, Alexander Mädche, and Christoph Spengel. How do tax compliance labels impact sharing platform consumers? an empirical study on the interplay of trust, moral, and intention to book. *Information Systems and e-Business Management*, 20(3):409–439, 2022.

11 Christian Engel, Edona Elshan, Philipp Ebel, and Jan M. Leimeister. Stairway to heaven or highway to hell: A model for assessing cognitive automation use cases. *Journal of Information Technology*, 39(1):94–122, 2024.

12 Peter A. François, Vincent Borghoff, Ralf Plattfaut, and Christian Janiesch. Why Companies Use RPA: A Critical Reflection of Goals. In *Business Process Management*, volume 13420, pages 399–417. 2022.

13 David Gefen, Elena Karahanna, and Detmar W Straub. Trust and TAM in online shopping: An integrated model. *MIS quarterly*, pages 51–90, 2003.

14 Jennifer Haase, Waldemar Kremser, Henrik Leopold, Jan Mendling, Linda Onnasch, and Ralf Plattfaut. Interdisciplinary Directions for Researching the Effects of Robotic Process Automation and Large Language Models on Business Processes. *Communications of the Association for Information Systems*, 54(1):579–604, 2024.

15 Lixiao Huang, Nancy J. Cooke, Robert S. Gutzwiller, Spring Berman, Erin K. Chiou, Mustafa Demir, and Wenlong Zhang. Chapter 13 – Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In *Trust in Human-Robot Interaction*, pages 301–319. Academic Press, 2021.

16 ISO/IEC. Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence, 2020.

17 Christian Janiesch, Marcus Fischer, Axel Winkelmann, and Valentin Nentwich. Specifying autonomy in the internet of things: the autonomy model and notation. *Inf. Syst. E Bus. Manag.*, 17(1):159–194, 2019.

18 CEN-CENELEC JTC. Architecture of standards, responding to european commission standardisation request on ai systems, 2021.

19 P. Madhavan and D. A. Wiegmann. Similarities and differences between human-human and human-automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, 8(4):277–301, 2007.

20 Riccardo Mariani, Francesca Rossi, Rita Cucchiara, et al. Trustworthy AI – part 1. *Computer*, 56(2):14–18, 2023.

21 Antonio Martínez-Rojas, J Sánchez-Oliva, José Manuel López-Carnicer, and Andrés Jiménez-Ramírez. AIRPA: An Architecture to Support the Execution and Maintenance of AI-Powered RPA Robots. In *Int. Conf. on Business Process Management*, pages 38–48, 2021.

22 Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *Academy of Management Review*, 20(3):709, 1995.

23 D Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2):1–25, 2011.

24 Carolyn McLeod. Trust. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2023 edition, 2023.

25 Jan Mendling, Gero Decker, Richard Hull, Hajo A. Reijers, and Ingo Weber. How do Machine Learning, Robotic Process Automation, and Blockchains Affect the Human Factor in Business Process Management? *Communications of the Association for Information Systems*, 43:297–320, 2018.

26 Artur Modliński, Damian Kedziora, Andrés Jiménez Ramírez, and Adela del Río-Ortega. Rolling Back to Manual Work: An Exploratory Research on Robotic Process Re-Manualization. In *Int. Conf. on Business Process Management*, pages 154–169, 2022.

27 Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. Part A*, 30(3):286–297, 2000.

28 EU Parliament. EU AI Act: First Regulation on Artificial Intelligence, 2024. `https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence`.

29 Simon Parsons, Katie Atkinson, Zimi Li, Peter McBurney, Elizabeth Sklar, Munindar Singh, Karen Haigh, Karl Levitt, and Jeff Rowe. Argument schemes for reasoning about trust. *Argument & Computation*, 5(2-3):160–190, 2014.

30 Rafael Cabello Ruiz, Andrés Jiménez-Ramírez, María José Escalona Cuaresma, and José González Enríquez. Hybridizing humans and robots: An RPA horizon envisaged from the trenches. *Comput. Ind.*, 138:103615, 2022.

31 Rehan Syed and Moe Thandar Wynn. How to Trust a Bot: An RPA User Perspective. In *Business Process Management: Blockchain and Robotic Process Automation Forum*, pages 147–160, 2020.

32 Magdalena Wischnewski, Nicole C. Krämer, and Emmanuel Müller. Measuring and Understanding Trust Calibrations for Automated Systems: A Survey of the State-Of-The-Art and Future Directions. In *2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023*, pages 755:1–755:16. ACM, 2023.

Participants

- Simone Agostinelli
Sapienza University of Rome, IT
- Marco Angelini
Link University – Rome, IT
- Aleksandre Asatiani
University of Gothenburg, SE
- Bernhard Axmann
TH Ingolstadt, DE
- Piercosma Bisconti
Interuniversity Consortium f. Computer Sci. – Rome, IT
- Angelo Casciani
Sapienza University of Rome, IT
- Christian Czarnecki
FH Aachen, DE
- Adela del Río Ortega
University of Sevilla, ES
- Andrea Delgado
University of the Republic – Montevideo, UY
- José González Enríquez
University of Sevilla, ES
- Glenda Hannibal
Paris Lodron Universität Salzburg, AT
- Christian Janiesch
TU Dortmund, DE
- Andrés Jiménez Ramírez
University of Sevilla, ES
- Faizan Ahmed Khan
University of Padova, IT
- Andrea Marrella
Sapienza University of Rome, IT
- Antonio Martínez Rojas
University of Sevilla, ES
- Artur Modlinski
University of Lodz, PL
- Ralf Plattfaut
Universität Duisburg-Essen, DE
- Jana-Rebecca Rehse
Universität Mannheim, DE
- Hajo A. Reijers
Utrecht University, NL
- Manuel Resinas
University of Sevilla, ES
- Michael Rosemann
Queensland University of Technology – Brisbane, AU
- Flávia Santoro
Inteli – São Paulo, BR
- Stefan Sarkadi
King's College London, GB
- Pnina Soffer
University of Haifa, IL
- Barbara Weber
Universität St. Gallen, CH
- Adriana Wilde
University of Southampton, GB