Seven Guidelines for Designing the User Interface in Robotic Process Automation

Judith Wewerka
Institute of Databases and Information Systems
Ulm University
Ulm, Germany
0000-0002-4809-2480

Christian Micus
Research and Development
BMW Group
Munich, Germany
0000-0002-1107-3964

Manfred Reichert
Institute of Databases and Information Systems
Ulm University
Ulm, Germany
0000-0003-2536-4153

***Abstract*—Robotic Process Automation (RPA) aims to automate rule-based business process tasks by software robots (bots) mimicking human interactions. Despite the partial automation achieved with RPA, humans still need to interact with the bots, which requires appropriate user interfaces. However, existing RPA research has not evaluated RPA from a software-ergonomic perspective so far and no corresponding user interface design guidelines exist. The objective of this paper is to evaluate the usability of RPA bots in industry and to provide user interface design guidelines to bot developers. The results we obtained from 50 questionnaires filled by RPA users indicate that both the input/output and the dialogue interfaces of RPA need to be improved, especially regarding *error tolerance*, *perceptibility*, *directability of user's attention*, *suitability for the task*, and *availability*. Finally, we derive seven guidelines for designing the user interface of RPA bots. Potential improvements include, among others, the quality of error messages, the efforts for error handling, and the monitoring of the current status of the tasks assigned to the bot.**

***Index Terms*—Robotic Process Automation, User Interface Design Guidelines, Software-Ergonomic Evaluation.**

I. INTRODUCTION

Robotic Process Automation (RPA) supports companies in optimizing and implementing business processes or parts of them with software robots (bots). The latter automate repetitive, rule-based process tasks and execute them in a human-like fashion [5], e.g., by transferring data from various input systems to system of records [26]. As opposed to general software systems, a bot is often implemented by domain experts (e.g., knowledge workers) rather than by professional software developers [41]. Note that RPA not only relieves humans from executing routine process tasks, but also changes their daily work procedures as the humans hand over tasks to bots [34]. Thus, RPA has a significant impact on the structure and organization of work activities, while changing the procedures and roles of the involved actors [34], [37].

To keep the human in the loop, [6] proposes a human-bot framework for RPA. However, a still unaddressed issue is how to realize the interactions between humans and RPA bots [40]. Even though RPA automates processes or parts of them, it does not provide an end-to-end automation and minimal interactions between the users and the bot are still required. These interactions include delegating tasks to the bot, reporting potential errors to the bot users, and informing the latter upon completion of a task (cf. Figure 1). In existing RPA implementations, the users often interact with the bot via e-mail, by storing documents in the filesystem of the bot and assigning the task to the bot in the respective system. The bot, in turn, interacts with the users via e-mail or by storing documents in a pre-defined filesystem [2], [14], [18], [40].

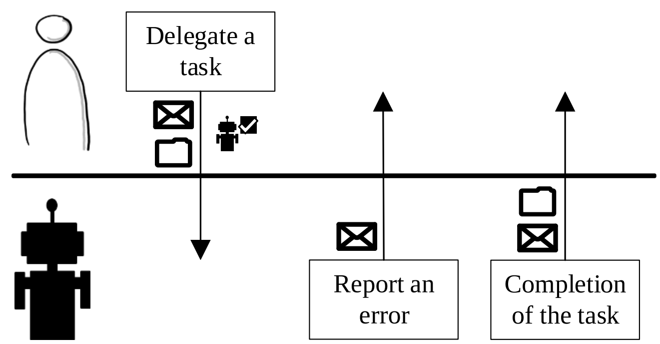

Fig. 1. Possible interactions between the users and the RPA bot.

A user-friendly design of the interactions between humans and RPA bot is needed to increase user acceptance [38]. However, as the knowledge workers implementing the RPA bots are usually no software developers [34], [43], they are inexperienced with optimizing the human bot interactions and, thus, demand for corresponding user interface design guidelines [40]. This paper evaluates seven RPA bots from automotive industry regarding their usability. Based on the evaluation results, we obtain seven guidelines for designing bot user interfaces, which shall help knowledge workers to optimize the bot user interface.

The remainder of the paper is organized as follows: Section II reviews literature on software ergonomics to obtain different procedures and criteria for assessing user interfaces. Section III adapts selected procedures and criteria to fit to the RPA context and to obtain a questionnaire for the software-ergonomic evaluation of a bot. The results we obtained from the distribution of the questionnaire to 400 employees in a knowledge-intensive domain are presented in Section IV. Finally, Section V derives seven guidelines for the RPA user interface design based on the results of the 50 filled questionnaires we obtained back.

II. Backgrounds

This section presents backgrounds on software ergonomics needed for understanding this work. Software Ergonomics intends to measure the usability of interactive software systems (software for short) [17]. Three aspects are considered: *adequate functionality*, i.e., the software efficiently supports user tasks, *correctness*, i.e., correspondence between actual software behavior and the predefined software performance, and *user interface*, i.e., the user's access to the software [30]. Figure 2 shows a user interface model (according to [16]), which actually distinguishes between three user interfaces [17]:

- The **input/output** interface covers the perceptibility and manageability of the software based on user input and software output.
- The **dialogue** interface comprises the interactions of the users with objects or functions of the software as well as their perception.
- The **tool** interface deals with the comfortable handling of the software and the access possibilities of the users to software functions.

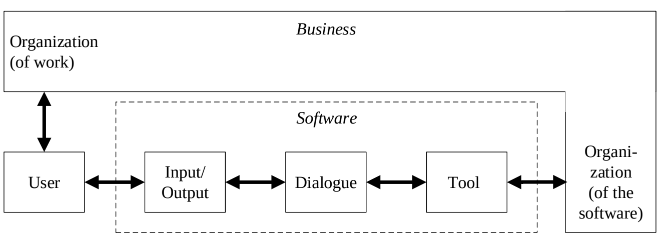

Fig. 2. A model of the user interfaces (adopted from [16])

Assessment Techniques. In general, there are three techniques to assess the usability of a software. The first one is *theory-based*, i.e., design principles are derived from theory and are evaluated in the software based on expert interviews and checklists. The main goal is to avoid fundamental weaknesses of the software [21]. The second technique is *task-based*, i.e., tasks are processed by the users to evaluate the usability and usefulness of the software [21]. The third technique is the *user-based* one, i.e., interviews, questionnaires, or experiments are conducted that ask users about expectations, impressions, and experiences with the software. The user-based method has a high informative value and is more objective than the other two methods [17], [21].

Procedures. As we focus on the interaction between humans and bot, we solely consider the assessment of the *user interfaces* with a user-based assessment technique. *Adequate functionality* and *correctness* are not considered. User-based techniques are chosen due to their objectivity and informative value. In the following, we present three procedures to assess the *user interfaces*:

- *ISO Norm 9241-110.* This norm sets the principles for designing the dialogue interface between humans and software. The aim is to examine drawbacks of the software based on the user's assessment of seven design requirements: suitability for the task, learnability, individualization, conformity with user's expectations, self-descriptiveness, controllability, and error tolerance [32].
- *IFIP user interface model.* The IFIP user interface model extends the ISO Norm with the organizational environment of the usage, i.e., the definition of rules determining the creation, definition, and distribution of work tasks [22]. The procedure systematically investigates four user interfaces: input/output, dialogue, tool, and organizational interface [16], [31]. The IFIP user interface model shall enable an objective assessment of the software and display complex behavior in a structured manner [22], [30].
- *IsoMetrics.* The model is built on the ISO Norm and provides 75 questions for the seven design requirements with a 5-point-likert answer scale. Additionally, each question is weighted by answering how important this aspect is for the overall impression of the software on a 5-point-likert scale [10], [12]. This rating allows the users to report problems with the software [12].

Assessment Criteria. The tool, dialogue, and input/output user interfaces are assessed based on different criteria. The *Tool Interface Criteria* are used to assess how comfortable the software can be handled as well as to assess access possibilities to the functional scope of the software [17], [30]. The criteria and explanations are given in Table I.

TABLE I
Tool Interface Criteria.

Criterion	Explanation
Availability	Software functions are available at any time.
Reliability	Software is reliable and does not generate system errors.
Reusability	Software can be used repeatedly and is deterministic.
Combinability	Users may combine functions.
Expandability	Users can program new functions.
Complexity	Operation of the software is easy.
Transparency	Function and response of the software to the user's input are predictable.

Dialogue Interface Criteria measure the interaction of the users with objects and functions of the software as well as the perception of the software by the users [9], [30]. The criteria and their explanations can be found in Table II.

Finally, *Input/Output Interface Criteria* evaluate the interaction based on the input of the users and output of the software [17], [30] (cf. Table III).

III. Questionnaire-based Evaluation of RPA Bot Usability

This section introduces the method we applied to evaluate the usability of RPA bots. In this context, we had slightly adapted the above presented procedures to fit to RPA bots. The resulting questionnaire is presented and its objectivity, validity, and reliability are assessed.

TABLE II
DIALOGUE INTERFACE CRITERIA.

Criterion	Explanation
Suitability for the task	Software supports the users in completing their process task.
Self-descriptiveness	For the users it is always obvious, which actions may be taken.
Conformity with user's expectations	Dialogue corresponds to the user's concerns.
Learnability	Software guides the users in learning how to use it.
Controllability	Users may start a dialogue and, then, influence its direction and speed until the user's goal is achieved.
Error tolerance	Despite incorrect entries, work results should be achievable with minimal correction efforts.
Individualization	Users may tailor the presentation of information to suit individual skills and needs.

TABLE III
INPUT/OUTPUT INTERFACE CRITERIA.

Criterion	Explanation
Perceptibility	The perceptibility of information, brightness, contrasts, and volume.
Legibility	Given by size, spacing, and line spacing of the software.
Distinctness	Information given by the software is clearly recognizable and distinguishable.
Clarity	Presentation and arrangement of information is precise.
Orientation support	The design of functional structures facilitates the user's orientation while using the software.
Directability of user's attention	Software can focus user's perception.
Manageability	Input systems of the software and corresponding feedback are useful.
Wholesomeness	Effect on users at their working place.

A. *RPA-specific Procedure*

To derive guidelines for designing the user interfaces of RPA bots, i.e., to meet our research goal, we combine the IFIP user interface and the IsoMetrics models (cf. Section II). The former investigates all user interfaces and, therefore, is suitable to account for the role changes of users coming with an RPA project. We exclude the organizational interface, which varies for different RPA bots and is not relevant for deriving user interface design guidelines. Moreover, the *weighting* aspect of the IsoMetrics model is included to address user needs and to obtain information about weaknesses of the RPA software [45]. All 14 criteria for evaluating the tool and dialogue interfaces (cf. Tables I and II) are considered. Concerning the 8 input/output interface criteria (cf. Table III), we exclude *wholesomeness* as it differs for each user and is not relevant in the context of our research goal.

For each of the remaining 21 criteria we consider two statements (i.e., 42 statements in total) that can be rated on a 6-point-likert scale (1-6). We choose an even scale to avoid a mediocre rating as feedback and to obtain a clear positive or negative tendency [29]. Moreover, each statement can be weighted as "unimportant" (1), "moderately important" (2), or "very important" (3) by participants depending on their satisfaction with the RPA bot they are working with. Table IV provides an excerpt from the questionnaire for the *availability* criterion.

The statements of all criteria for assessing the three user interfaces are summarized in Table V.[1] These statements are based on [17], [30], but have been slightly adapted to account for the peculiarities of the RPA software.

We invite 400 engineers of a global automotive vendor to participate in the survey. The subjects are selected based on the following criteria:

- The RPA bot takes over a task that has been accomplished by the engineer before. Hence, the participants are RPA users.
- The interaction between the RPA bot and the engineer follows the schema depicted in Figure 1.

B. *Objectivity, Validity, and Reliability of the Questionnaire*

Before presenting the results obtained with the questionnaire, we assess its objectivity, validity, and reliability. In general, the IsoMetrics questionnaire is considered as reliable and valid [12]. However, as we slightly adopted this questionnaire, we need to look at general measures to ascertain its objectivity, validity, and reliability. A confirmatory factor analysis is performed to test whether the statements measuring the criteria are consistent with our understanding of the criteria [20]. Structural equation modeling is used with the *lavaan* package in R. Figure 3 shows the aspects to be assessed and the measures used (in *italic*).

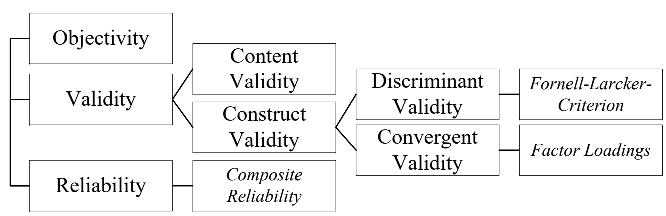

Fig. 3. Overview of objectivity, validity, and reliability.

Objectivity defines the extent to which results are independent from the respective respondent [8]. It is achieved through the design of the questionnaire, which is the same for every participant, and the use of rating scales [10]. Moreover, the evaluation of the scores needs to be standardized and easy to understand [10]. All aspects are covered by the questionnaire, i.e., objectivity is given.

Validity assesses the extent to which a criterion is accurately measured by the statements [15]. To ensure validity, content and construct validity need to be considered [10]. **Content validity** is given if the statements adequately cover the meaning of the variable to be assessed [15]. The questionnaire follows the IFIP user interface as well as the IsoMetrics

[1]The original questionnaire as well as the raw data of the 50 returned questionnaire instances are available via the following Link to Researchgate

TABLE IV
EXCERPT FROM THE QUESTIONNAIRE: *Availability* CRITERION.

Availability Criterion	*Rating:* Please indicate your consent to this statement.	*Weighting:* Please weight the importance of this statement regarding your satisfaction with the RPA bot.
Statement 1: The RPA bot is always available when I want to use it.	☐ I strongly disagree ☐ I moderately disagree ☐ I somewhat disagree ☐ I somewhat agree ☐ I moderately agree ☐ I strongly agree	☐ Unimportant ☐ Moderately important ☐ Very important
Statement 2: My work with the RPA bot is not affected by disruptions or long response times.	☐ I strongly disagree ☐ I moderately disagree ☐ I somewhat disagree ☐ I somewhat agree ☐ I moderately agree ☐ I strongly agree	☐ Unimportant ☐ Moderately important ☐ Very important

models. Both are based on ISO Norm 9241-110 and use two statements for each criterion. Consequently, content validity is ensured. **Construct validity**, in turn, refers to the degree to which the score of a statement corresponds to the criterion the measure is intended to operationalize [11]. It is further divided into discriminant and convergent validity [10]. **Discriminant validity** corresponds to the degree to which a criterion is truly distinct from others. The *Fornell-Larcker-Criterion* is used to assess discriminant validity by assessing whether each criterion shares more variance with the corresponding statement than with other statements in the questionnaire [13]. Therefore, the diagonal elements in Table VIII, which are the square roots of the average variance extracted, must be greater than the correlations of the latent variables in the off-diagonal elements [1]. This is fulfilled for all criteria and discriminant validity is given. Finally, **convergent validity** is evaluated. Here, we assess whether the statements measuring a criterion behave as if they are measuring a common underlying variable [7]. *Factor Loadings* corresponding to the correlation coefficients of each statement with its criterion are used for evaluation [4]. Values range from 0.71 (*MAN 2*) to 0.99 (*SUI 1*) (cf. Table VI) and exceed the suggested threshold of 0.70 [28]. Hence, convergent validity is fulfilled and validity in general is given for our questionnaire.

Reliability refers to the accuracy of the questionnaire [15]. We use *composite reliability* to measure internal consistency [3] and obtain values between 0.69 (*legibility*) and 0.97 (*suitability for the task*) (cf. Table VI). All values exceed the threshold of 0.60 as suggested for exploratory research [3].

IV. RESULTS

This section evaluates the questionnaires returned back by the participants. First, results of the **descriptive analysis** are given. Second, the rating and weight of the statements is detailed. We receive 50 of the 400 questionnaires back, resulting in a response rate of 12.5%. The respondents are working with seven different RPA bots in total that realize the human robot interaction as shown in Figure 1. All seven bots have been released in 2020 and can be seen as state-of-the-art bots. Regarding the experience with RPA (cf. Figure 4a), 16% (N=8) of the respondents have been using the bot less than a month, 36% (N=18) between one and three months, 14% (N=7) between three and six months, and 34% (N=17) for more than six months. Regarding age (cf. Figure 4b): 10% (N=5) of the respondents are 30 years or younger, 56% (N=28) are between 31 and 40 years old, 26% (N=13) between 41 and 50, and 8% (N=4) older than 50.

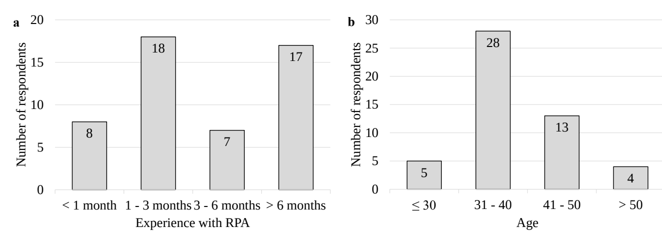

Fig. 4. Descriptive Statistics: a) Experience with RPA, b) Age.

In the first step, we evaluate the 21 criteria based on the average **rating and weight of the statements** in the questionnaire, i.e., the answer scores corresponding to the two statements of a criterion are averaged (cf. Table VII). Note that the rating is based on a 6-point-likert scale and the weighting on a 3-point-likert scale. The values for rating the criteria lie between 3.32 for *individualization* and 4.99 for *expandability*. The top five rated criteria are *expandability* (4.99), *orientation support* (4.97), *reusability* (4.95), *complexity* (4.90), and *conformity with user's expectation* (4.84). The weights, which indicate the importance of the criteria, lie between 2.00 for *individualization* and 2.85 for *reliability* and *suitability for the task*. The criteria evaluated with the five highest weights are *reliability* and *suitability for the task* (2.85), *directability of user's attention* (2.74), *conformity with user's expectations* (2.67), and *expandability* (2.66).

In the second step, we evaluate the three user interfaces. For this purpose, we take the median of the data from the criteria to the three interfaces (cf. medians in Table VII). The

TABLE V
STATEMENTS FOR THE THREE INTERFACES.

Tool interface	Availability	*AVA 1:* The RPA bot is always available when I want to use it. *AVA 2:* My work with the RPA bot is not affected by disruptions or long response times.
	Reliability	*REL 1:* The RPA bot works as required without any complications. *REL 2:* When I work with the RPA bot, no system errors (e.g. crash or incorrect execution) occur.
	Reusability	*REU 1:* I can run the RPA bot as often as I want. *REU 2:* If I use the RPA bot repeatedly, I get the same result with the same input, i.e., the RPA bot behaves deterministically.
	Combinability	*COM 1:* The RPA bot has a modular structure and can be used for sub-tasks as well. *COM 2:* It is easy to use the RPA bot for similar tasks.
	Expandability	*EXP 1:* If the system or the task changes, the RPA bot can be adapted. *EXP 2:* The RPA bot can be expanded to include additional sub-tasks.
	Complexity	*COP 1:* The terms the RPA bot uses are understandable to me. *COP 2:* It is easy to run the RPA bot to handle the task as desired.
	Transparency	*TRA 1:* The result of the RPA bot is predictable for me. *TRA 2:* The RPA bot gives me feedback on the progress of my task.
Dialogue interface	Suitability for the task	*SUI 1:* The RPA bot is well aligned to meet the requirements of my working tasks. *SUI 2:* The RPA bot supports me in completing my tasks and is not an additional burden.
	Self-descriptiveness	*SED 1:* The RPA bot gives me enough information about what inputs are allowed and what data may be used. *SED 2:* I am fully aware of the purpose and scope of the RPA bot.
	Conformity with user's expectations	*CON 1:* The RPA bot works exactly as I expect it to. *CON 2:* The RPA bot performs the task in the same way as if done manually.
	Learnability	*LEA 1:* I only have to remember few details to run the RPA bot. *LEA 2:* The RPA bot requires little learning and supports me in learning how to use it.
	Controllability	*CTR 1:* I can adapt the type and scope of RPA inputs and outputs (e.g., the result is available in different file formats). *CTR 2:* I can adapt the reaction time and the speed of executing the RPA bot to my individual needs.
	Error tolerance	*ERR 1:* The RPA bot creates easily understandable error messages that help me to fix the error. *ERR 2:* The RPA bot only requires little efforts for correcting errors.
	Individualization	*IND 1:* The workflow and the processing order of the RPA bot can be adjusted to my individual needs. *IND 2:* The RPA bot can be easily adapted to my personal way of working.
Input/output interface	Perceptibility	*PER 1:* The information required for processing tasks is always in the right place on the screen. *PER 2:* It can be seen whether the RPA bot has completed the task.
	Legibility	*LEG 1:* The readability of the texts and characters created by the RPA bot is good. *LEG 2:* The results of the RPA bot can be presented according to my individual requirements.
	Distinctness	*DIS 1:* I can clearly assign the feedback received from the RPA bot to the triggering process. *DIS 2:* Results the RPA bot delivers cannot be distinguished from the ones obtained through manual processing.
	Clarity	*CLA 1:* Information and messages from the RPA bot are clearly displayed. *CLA 2:* Information and messages from the RPA bot are displayed in the same way on different output media.
	Orientation support	*ORS 1:* The good design of the RPA bot eases its use. *ORS 2:* The representations of the RPA bot are consistent and foster its use.
	Directability of user's attention	*DOA 1:* The RPA bot clearly indicates when a task is completed, a task is aborted, or problems occur. *DOA 2:* The RPA bot does not stop me from doing other work while it is running.
	Manageability	*MAN 1:* The RPA bot can be operated individually, not just following a rigid procedure. *MAN 2:* If the RPA bot or I make a mistake during task processing, I can easily undo the faulty operation and restore the original data.

tool interface is rated with 4.74 and weighted with 2.55, the **dialogue** interface has a rating of 4.43 and a weight of 2.53, and the **input/output** interface is evaluated with 4.37 and weighted with 2.47. Note that we decided to use the median instead of the average value, as it is more robust to outliers and better suited for ordinal scales [19].

In the third step, we evaluate the relationship between rating and weight (cf. Figure 5). The distance between these two values provides information on the acceptance of the criteria [12]. We use a bubble chart to illustrate this relationship [36]. The chart is divided along the two median values of all criteria (i.e., 4.53 for the rating and 2.52 for the weight) into four parts:

- Top right: criteria with a *high rating being important* for the users.
- Bottom left: criteria with a *low rating not being important* for the users.
- Top left: criteria with a *low rating*, but *being important* for the users - these criteria should be put more into focus.
- Bottom right: criteria with a *high rating*, but *not being important* for the users - these criteria should be less in the focus.

The criteria having a low rating and being weighted as important include *error tolerance*, *perceptibility*, *directability of user's attention*, *availability*, and *suitability for the task*. The latter lies on the line of the median and is, therefore, considered as well.

Contrary, the criteria with high rating and weighted as unimportant are *distinctness*, *legibility*, and *transparency*. We add *reusability* and *orientation support*, which both lie on the median line, to this list.

V. DERIVING GUIDELINES FOR RPA USER INTERFACE DESIGN

Based on the results presented in Section IV, this section derives seven guidelines for designing the user interfaces of RPA bots. These guidelines shall help knowledge workers without IT background to successfully implement RPA bots

TABLE VI
MEAN, FACTOR LOADING, AND COMPOSITE RELIABILITY FOR EACH STATEMENT.

Criterion	Statement	Mean	Factor Loading	Composite Reliability
Availability	AVA 1/AVA 2	4.57/4.33	0.82/0.76	0.77
Reliability	REL 1/REL 2	4.86/4.60	0.89/0.81	0.84
Reusability	REU 1/REU 2	4.57/5.33	0.78/0.76	0.75
Combinability	COM 1/COM 2	4.25/4.31	0.76/0.75	0.73
Expandability	EXP 1/EXP 2	5.00/4.97	0.94/0.84	0.89
Complexity	COP 1/COP 2	5.12/4.68	0.79/0.76	0.75
Transparency	TRA 1/TRA 2	5.12/4.36	0.78/0.75	0.74
Suitability for the task	SUI 1/SUI 2	4.63/4.43	0.99/0.94	0.97
Self-descriptiveness	SED 1/SED 2	4.29/4.93	0.75/0.85	0.78
Conformity with user's expectations	CON 1/CON 2	4.85/4.83	0.91/0.90	0.90
Learnability	LEA 1/LEA 2	4.51/4.35	0.88/0.94	0.91
Controllability	CTR 1/CTR 2	3.94/3.11	0.80/0.96	0.87
Error tolerance	ERR 1/ERR 2	3.54/4.06	0.91/0.83	0.87
Individualization	IND 1/IND 2	3.08/3.56	0.82/0.79	0.79
Perceptibility	PER 1/PER 2	4.09/4.39	0.78/0.79	0.77
Legibility	LEG 1/LEG 2	4.92/4.31	0.74/0.72	0.69
Distinctness	DIS 1/DIS 2	4.88/4.20	0.82/0.85	0.82
Clarity	CLA 1/CLA 2	4.53/3.97	0.82/0.90	0.85
Orientation support	ORS 1/ORS 2	4.92/5.03	0.82/0.81	0.80
Directability of user's attention	DOA 1/DOA 2	3.93/4.82	0.79/0.83	0.79
Manageability	MAN 1/MAN 2	3.63/4.26	0.80/0.71	0.73

TABLE VII
RESULTS OF THE SOFTWARE ERGONOMIC EVALUATION.

Interface	Criterion	Rating	Weight
Tool	Availability	4.45	2.58
	Reliability	4.73	2.85
	Reusability	4.95	2.52
	Combinability	4.28	2.36
	Expandability	4.99	2.66
	Complexity	4.90	2.55
	Transparency	4.74	2.41
Tool - Median		**4.74**	**2.55**
Dialogue	Suitability for the task	4.53	2.85
	Self-descriptiveness	4.61	2.53
	Conformity with user's expectations	4.84	2.67
	Learnability	4.43	2.27
	Controllability	3.53	2.08
	Error tolerance	3.80	2.59
	Individualization	3.32	2.00
Dialogue - Median		**4.43**	**2.53**
Input/ Output	Perceptibility	4.24	2.59
	Legibility	4.62	2.47
	Distinctness	4.54	2.33
	Clarity	4.25	2.19
	Orientation support	4.97	2.52
	Directability of user's attention	4.37	2.74
	Manageability	3.95	2.30
Input/Output - Median		**4.37**	**2.47**
Median		**4.53**	**2.52**

satisfying the needs of the users in respect to their interaction with the RPA bot.

Five of the top seven rated criteria refer to the tool interface. Contemporary RPA implementations have focused on implementing reliable, reusable, expandable, simple, and transparent RPA bots [25], [26]. Consequently, these criteria are now the ones with the highest evaluation. When taking a look at the criteria with the highest weights, i.e., the highest importance for the users, there are criteria from all three user interfaces: Of the top eight weighted criteria, three refer to the tool, three to the dialogue, and two to the input/output user interfaces (cf. Table VII). Due to the discrepancy between importance and rating of the criteria, RPA bot developers need not only focus on reliable and expandable implementations, but also on aspects of the dialogue and input/output interfaces. Among others, these include the *suitability for the task, directability of user's attention, error tolerance,* and *perceptibility*.

Examining the median evaluation values of the three user interfaces, the **tool** interface is the best evaluated one (4.74) and is weighted as most important (2.55). Note that the weighting of the **dialogue** (2.53) and **input/output** interfaces (2.47) are nearly as high. The ratings of the dialogue (4.43) and the input/output interfaces (4.37), however, are far below the one of the tool interface and, therefore, should be put more into focus.

Seven Guidelines for Designing the User Interface in Robotic Process Automation

1. Improve the quality and comprehensibility of the bot error messages.
2. Minimize the efforts for correcting bot errors.
3. Ensure visibility of the current status of the task.
4. Attract the attention of users when their task is completed, aborted, or any problem occurs.
5. Guarantee that the users obtain the results form the bot within a reasonable response time.
6. Take care that no additional efforts are required to use the RPA bot.
7. Do not over-emphasize legibility, transparency, and distinctness.

We derive the guidelines by investigating the criteria that show a discrepancy between rating and weight. We emphasize

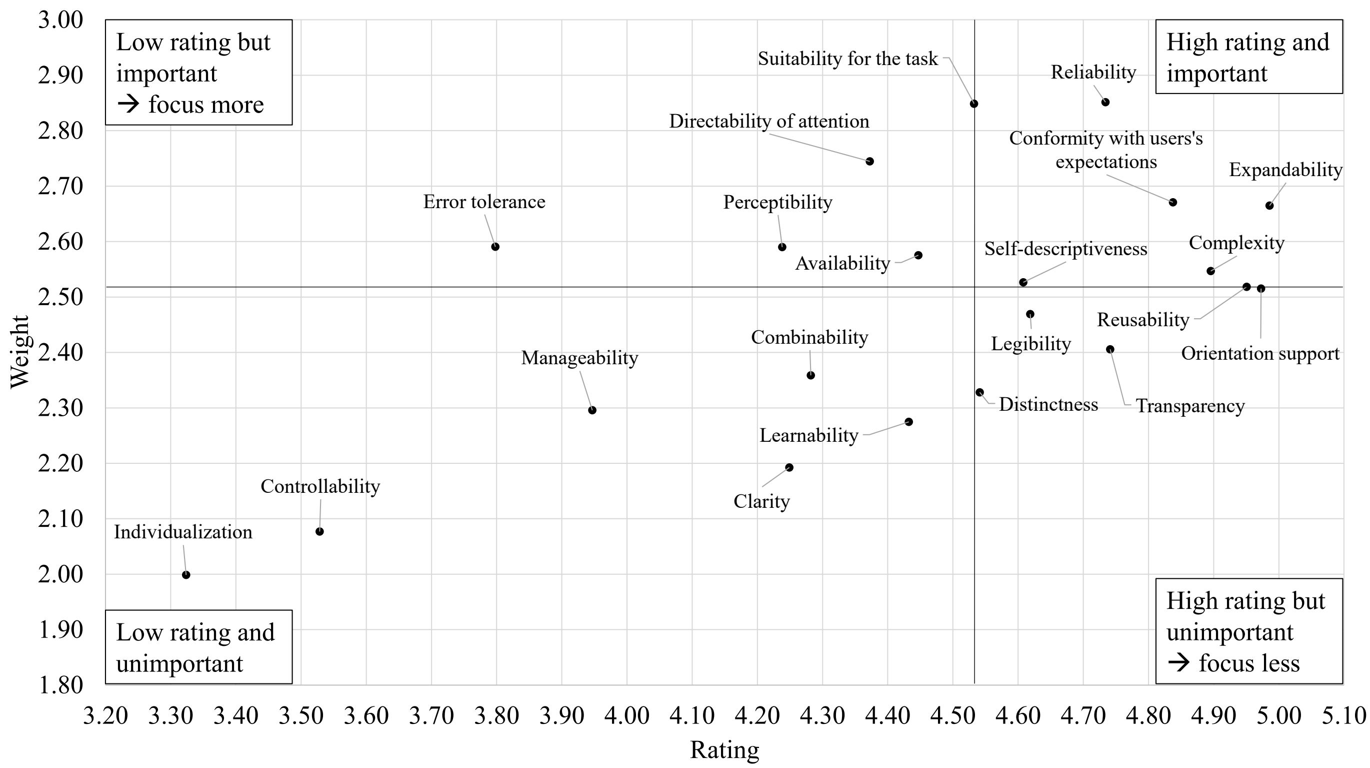

Fig. 5. Relationship between rating and weight of the criteria.

five criteria with low rating and high weight. Note that four of them refer to the dialogue or the input/output interfaces. Possible improvements of the interaction between human and bot are indicated in Figure 6.

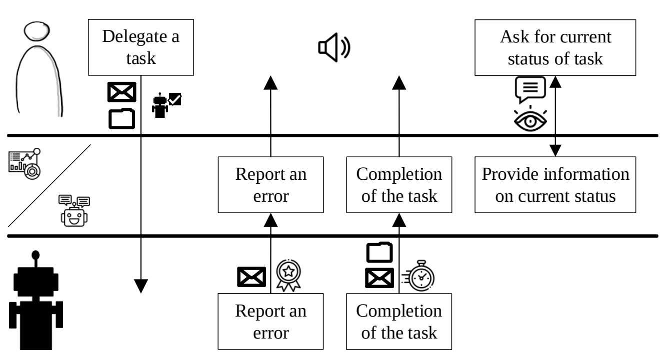

Fig. 6. Improved interactions between the users and the RPA bot.

First and foremost, *error tolerance* needs to be ensured. Both statements related to this criterion are poorly rated. As a consequence, developers should **improve the quality and comprehensibility of the bot error messages (1)** to support users in understanding and fixing the errors (cf. quality symbol in Figure 6). In a second step, the **efforts for correcting bot errors need to be minimized (2)**, i.e., the users should be able to solve errors and problems without being an RPA expert. Note that this aspect is rarely addressed in literature, only [18] suggests creating novel solutions for error handling.

Another criterion to which more attention should be paid during bot implementation is *perceptibility*. According to the results, users complain that it does not become transparent when a bot has completed an assigned task. Consequently, the RPA input/output interface needs to be improved. A simple notification via e-mail seems to be not sufficient. Instead, a chatbot monitoring the progress and answering questions like "How far is the processing of task xy?" or "Can you send me a chat message when task xy is finished?" could **ensure the visibility of the current status of the task (3)**. Alternatively, a dashboard showing all tasks assigned to the bot as well as their status, e.g., "being processed" or "in queue", can be used (see the layer between human and bot in Figure 6). Then, users can monitor whether their tasks are completed or assess on their own how long they have to wait for completion taking the number of tasks in the bot queue into account (see the interaction between users and dashboard/chatbot in Figure 6 on the right). In literature, the combination of RPA bots and chatbots is at its beginnings. [33] develops a framework for proactive chatbots communicating with the users. Their aim is to provide a user-friendly interface to spread RPA adoption. Using a dashboard to report the status of RPA bots is proposed in [23], [35].

The low-rated criterion *directability of user's attention* goes into the same direction. The first statement, i.e., the bot clearly indicates when a task is completed, aborted, or problems occur, is poorly assessed. One could think of different possibilities to **attract the attention of the users (4)**, e.g., popping-up of a text message or playing a sound (cf. sound symbol in Figure 6). The bot should not only send an e-mail to the users upon completion, but also inform them if any error occurs. So far, literature has focused on communicating task completion to the users by e-mail [2], [14], [18], [40]. No other types of communication are reported.

Regarding the *availability* criterion, its second statement is poorly rated. During RPA implementation, therefore, it needs to be guaranteed that users **obtain results from the bot within a reasonable response time (5)** (cf. speed symbol in Figure 6). In general, availability is assumed to be improved by RPA. Several publications emphasize that RPA bots are 24/7 available [24], [27]. However, improving processing time with RPA must not be taken for granted [39]. Some case studies achieve fast RPA bots, e.g., minutes instead of days [42] or 15 minutes instead of 6 hours [44]. Other projects do not improve response times, e.g., 431 seconds instead of 440 seconds [2]. However, the feeling to obtain the results in a reasonable amount of time remains subjective.

Based on the described guidelines, one can assume that the *suitability for the task* can be improved as well. Currently, users complain that the RPA bot introduces additional efforts. If, on the contrary, users are informed about the status of their task (perceptibility) or are immediately informed upon task completion (directability of user's attention) or are provided with useful error messages to correct errors themselves (error tolerance), and the bot is providing answers within a reasonable response time (availability), the RPA image should improve and it might be seen as a real support. Therefore, the RPA developer should **take care that no additional efforts are required to use the RPA bot (6).**

Finally, concerning the criteria with high rating and low weight, we recommend that the developers **no longer overemphasize those aspects (7)**, e.g., legibility or transparency. Obviously, a transparent software providing distinguishable and legible information is important to its users, but the main focus needs to be shifted.

VI. Conclusions

This work focuses on the design of the user interface for RPA bots, which is subdivided into the tool, dialogue, and input/output interfaces. 50 engineers from an automotive company, who are experienced in interacting with at least one RPA bot, participated in the survey. The latter asks for ratings and weights of 21 different criteria. The survey confirms that the tool interface of contemporary bots is well perceived by users. By contrast, the dialogue and input/output interfaces for RPA need to be improved. Finally, we derive seven guidelines for designing user interfaces in RPA. In future work, the survey needs to be repeated in other domains to ensure generalizability of results. The evaluation can help companies to implement RPA more successfully by optimizing the user interface design. The derived guidelines should be followed and monitored whether the rating of the RPA software improves.

Appendix

References

[1] M. Ab Hamid, W. Sami, and M. Sidek, "Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion," in *J Physics Conf S*, vol. 890, 2017.

[2] S. Aguirre and A. Rodriguez, "Automation of a Business Process Using Robotic Process Automation (RPA): A Case Study," in *Works Eng Appl*, 2017, pp. 65–71.

[3] M. Al-Emran, V. Mezhuyev, and A. Kamaludin, "PLS-SEM in Information Systems Research: A Comprehensive Methodological Reference," in *Int Conf Adv Intell Syst Informat*, 2018, pp. 644–653.

[4] J. C. Anderson and D. W. Gerbing, "Structural equation modeling in practice: A review and recommended two-step approach," *Psychol Bull*, vol. 103, no. 3, pp. 411–423, 1988.

[5] A. Asatiani and E. Penttinen, "Turning robotic process automation into commercial success - Case OpusCapita," *J Inform Technol Teach Cases*, vol. 6, no. 2, pp. 67–74, 2016.

[6] R. Cabello, M. J. Escalona, and J. G. Enríquez, "Beyond the Hype: RPA Horizon for Robot-Human Interaction," in *Int Conf Bus Proc Manag*, 2020, pp. 185–199.

[7] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q*, vol. 13, no. 3, pp. 319–339, 1989.

[8] N. Döring and J. Bortz, *Forschungsmethoden und Evaluation*. Springer, 2016.

[9] W. Dzida, "The Development of Ergonomic Standards," *SIGCHI Bull*, vol. 20, no. 3, pp. 35–42, 1989.

[10] K. Figl, "Deutschsprachige Fragebögen zur Usability-Evaluation im Vergleich," *Zeits Arbeitswissen*, vol. 4, pp. 321–337, 2010.

[11] M. Fishbein and I. Ajzen, *Belief, attitude, intention, and behavior: An introduction to theory and research*, 1977.

[12] G. Gediga, K.-C. Hamborg, and I. Düntsch, "The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems," *Behav Inform Technol*, vol. 18, no. 3, pp. 151–164, 1999.

[13] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a Silver Bullet," *J Market Theo Prac*, vol. 19, no. 2, pp. 139–152, 2011.

[14] P. Hallikainen, R. Bekkhus, and S. L. Pan, "How OpusCapita Used Internal RPA Capabilities to Offer Services to Clients," *MIS Q Exec*, vol. 17, no. 1, pp. 41–52, 2018.

[15] R. Heale and A. Twycross, "Validity and reliability in quantitative studies," *Evid Nurs*, vol. 18, no. 3, pp. 66–67, 2015.

[16] A. M. Heinecke, "Software ergonomics for real-time systems," in *Human Computer Interaction*, T. Grechenig and M. Tscheligi, Eds., 1993, pp. 377–390.

[17] M. Herczeg, *Software-Ergonomie: Theorien, Modelle und Krierien für gebrauchstaugliche interaktive Computersysteme*. Walter de Gruyter GmbH & Co KG, 2018.

[18] J. Hindel, L. M. Cabrera, and M. Stierle, "Robotic process automation: Hype or hope?" in *15th Int Conf Wirtschaftsinformatik*, 2020.

[19] S. Jamieson, "Likert scales: How to (ab)use them?" *Med educ*, vol. 38, no. 12, pp. 1217–1218, 2004.

[20] K. G. Jöreskog, "A general approach to confirmatory maximum likelihood factor analysis," *Psychometrika*, vol. 34, no. 2, pp. 183–202, 1969.

[21] J. Karat, "Software Evaluation Methodologies," in *Handbook of Human-Computer Interaction*, M. Helander, Ed., 1988, pp. 891–903.

[22] M. Koch, H. Reiterer, and A. M. Tjoa, "Kriterien zur Gestaltung und Bewertung menschengerechter Arbeit," in *Softw Ergonom*, 1991, pp. 43–86.

[23] J. Kokina and S. Blanchette, "Early evidence of digital labor in accounting: Innovation with Robotic Process Automation," *Int J Acc Inform Syst*, vol. 35, 2019.

[24] M. Lacity and L. Willcocks, "A New Approach to Automating Services," *MIT Sloan Manag Rev*, 2017.

[25] M. Lacity, L. Willcocks, and A. Craig, "Robotic Process Automation: Mature Capabilities in the Energy Sector," *Outs Unit Work Res Pap S*, pp. 1–19, 2015.

[26] ——, "Robotizing Global Financial Shared Services at Royal DSM," *Outs Unit Work Res Pap S*, vol. 16, no. 2, pp. 1–26, 2016.

[27] ——, "Service Automation: Cognitive Virtual Agents at SEB Bank," *London School Econ Polit Sci*, pp. 1–29, 2017.

TABLE VIII
FORNELL-LARCKER-CRITERION.

	AVA	REL	REU	COM	EXP	COP	TRA	SUI	SED	CON	LEA	CTR	ERR	IND	PER	LEG	DIS	CLA	ORS	DOA	MAN
AVA	**0.79**																				
REL	0.70	**0.87**																			
REU	0.70	0.68	**0.77**																		
COM	0.53	0.41	0.54	**0.76**																	
EXP	0.36	0.50	0.52	0.61	**0.89**																
COP	0.53	0.50	0.75	0.61	0.49	**0.80**															
TRA	0.49	0.44	0.40	0.61	0.45	0.35	**0.77**														
SUI	0.70	0.31	0.39	0.47	0.21	0.26	0.24	**0.97**													
SED	0.58	0.26	0.31	0.64	0.26	0.15	0.38	0.74	**0.81**												
CON	0.66	0.26	0.40	0.47	0.12	0.20	0.44	0.91	0.77	**0.91**											
LEA	0.54	0.14	0.37	0.60	0.08	0.47	0.31	0.65	0.67	0.72	**0.91**										
CTR	0.68	0.32	0.41	0.69	0.22	0.49	0.33	0.63	0.69	0.59	0.83	**0.89**									
ERR	0.55	0.38	0.48	0.73	0.58	0.39	0.56	0.49	0.63	0.52	0.58	0.74	**0.87**								
IND	0.53	0.28	0.28	0.75	0.30	0.33	0.37	0.71	0.78	0.68	0.63	0.79	0.75	**0.83**							
PER	0.64	0.41	0.42	0.68	0.33	0.47	0.43	0.65	0.67	0.67	0.78	0.75	0.67	0.68	**0.79**						
LEG	0.36	0.24	0.34	0.61	0.29	0.37	0.07	0.60	0.72	0.46	0.66	0.66	0.52	0.62	0.66	**0.74**					
DIS	0.27	0.09	0.28	0.67	0.37	0.18	0.27	0.61	0.68	0.63	0.54	0.45	0.47	0.63	0.47	0.59	**0.87**				
CLA	0.46	0.32	0.35	0.76	0.37	0.31	0.38	0.58	0.83	0.59	0.68	0.82	0.75	0.84	0.65	0.77	0.72	**0.86**			
ORS	0.56	0.19	0.38	0.60	0.27	0.39	0.18	0.73	0.66	0.67	0.77	0.77	0.70	0.77	0.74	0.65	0.70	0.73	**0.82**		
DOA	0.53	0.15	0.46	0.35	0.10	0.29	0.09	0.81	0.45	0.75	0.64	0.49	0.39	0.46	0.57	0.53	0.60	0.41	0.77	**0.85**	
MAN	0.52	0.18	0.32	0.43	0.22	0.31	0.29	0.49	0.41	0.57	0.61	0.66	0.64	0.59	0.47	0.19	0.48	0.52	0.71	0.54	**0.75**

[28] V. Mezhuyev, M. Al-Emran, M. A. Ismail, L. Benedicenti, and D. A. Chandran, "The Acceptance of Search-Based Software Engineering Techniques: An Empirical Evaluation Using the Technology Acceptance Model," *IEEE Access*, vol. 7, 2019.

[29] T. Nemoto and D. Beglar, "Likert-scale questionnaires," in *JALT Conf Proceed*, 2014, pp. 1–8.

[30] R. Oppermann, B. Murchner, and M. Koch, *Software-ergonomische Evaluation: Der Leitfaden EVADIS II*. Berlin [ua]: de Gruyter, 1992.

[31] B. Preim, *Entwicklung Interaktiver Systeme: Grundlagen, Fallbeispiele und innovative Anwendungsfelder*. Springer-Verlag, 1999.

[32] J. Prümper, "Methode Isonorm-Fragebogen," *B Wirt Energ*, pp. 1–6, 2008.

[33] Y. Rizk, A. Bhandwalder, S. Boag, T. Chakraborti, V. Isahagian, Y. Khazaeni, F. Pollock, and M. Unuvar, "A Unified Conversational Assistant Framework for Business Process Automation," *arXiv:2001.03543*, 2020.

[34] C. Rutschi and J. Dibbern, "Towards a framework of implementing software robots: Transforming Human-executed Routines into Machines," *Data Base Adv Inform Syst*, vol. 51, no. 1, pp. 104–128, 2020.

[35] M. Schmitz, C. Dietze, and C. Czarnecki, "Enabling digital transformation through robotic process automation at Deutsche Telekom," in *Digitalization Cases, Management for Professionals*, N. Urbach and M. Röglinger, Eds., 2019, pp. 15–33.

[36] S. Sirisack and A. Grimvall, "Visual detection of change points and trends using animated bubble charts," *Environ Monitor*, pp. 327–340, 2011.

[37] J. Wanner, A. Hofmann, M. Fischer, F. Imgrund, C. Janiesch, and J. Geyer-Klingeberg, "Process selection in RPA projects - Towards a quantifiable method of decision making," *40th Int Conf Inform Syst*, pp. 1–17, 2020.

[38] J. Wewerka, S. Dax, and M. Reichert, "A User Acceptance Model for Robotic Process Automation," in *24th Int Ent Dist Obj Comp Conf*, 2020, pp. 97–106.

[39] J. Wewerka and M. Reichert, "Towards Quantifying the Effects of Robotic Process Automation," in *24th Int Ent Dist Obj Comp Conf Works*, 2020, pp. 11–19.

[40] ——, "Robotic Process Automation in the Automotive Industry - Lessons Learned form an Exploratory Case Study," in *15th Int Conf Res Chall Inform Sci*, 2021, pp. 1–17.

[41] L. Willcocks and M. Lacity, "Robotic Process Automation: The Next Transformation Lever for Shared Services," *Outs Unit Work Res Pap S*, vol. 15, no. 7, pp. 1–35, 2015.

[42] ——, "Robotic Process Automation at Telefónica O2," *MIS Q Exec*, vol. 15, no. 1, pp. 21–35, 2016.

[43] L. Willcocks, M. Lacity, and A. Craig, "The IT Function and Robotic Process Automation," *Outsourc Unit Work Res Pap S*, vol. 15, no. 5, pp. 1–39, 2015.

[44] W. William and L. William, "Improving Corporate Secretary Productivity using Robotic Process Automation," *Int Conf Technol Appl AI*, 2019.

[45] R. C. Williges, B. H. Williges, and J. Elkerton, "Software interface design," in *Handbook of human factors*, 1987, pp. 1416–1449.