TOOLS

Chatbots in psychological screening: A report on our experiences

Roman Soucek and Amanda S. Voss

ROMAN SOUCEK

Department of Psychology, MSH Medical School Hamburg, Germany

roman.soucek@medicalschool-hamburg.de

AMANDA S. VOSS

Institute and Outpatient Clinic of Occupational, Social and Environmental Medicine, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

About the authors

Roman Soucek is a Full Professor of Work and Organizational Psychology at the Department of Psychology at the MSH Medical School Hamburg in Germany. His research interests include resilience in the work context, work intensity, human resource development, and the evaluation of training interventions. He is also an experienced freelance trainer with a particular interest in bridging the gap between science and practice.

Amanda S. Voss is a research associate at the Institute and Outpatient Clinic of Occupational, Social and Environmental Medicine at the Friedrich-Alexander-Universität Erlangen-Nürnberg. As a sociologist, she is part of the work group on public health which examines physical and psychological stress at work and its influence on the health of employees. Amanda's research focuses on mental health at work, work intensity and workplace health management.

Abstract

The rapid development of chatbots powered by artificial intelligence has opened up a wide range of applications in psychology. For industrial and organizational psychology, chatbots are an interesting area of research, as they are increasingly accepted and integrated into the work context. They also offer a rich potential as a research tool, especially in the areas of diagnostics and intervention. In this article, we will discuss the use of chatbots in the context of psychological screening. First, we will describe how chatbots work as conversational agents and give a brief overview of the use of chatbots in psychological assessment. We then describe our own approach to developing a chatbot to assess different facets of work intensity. Finally, we discuss our experiences and perspectives on the use of chatbots in psychological screening.

Keywords: chatbots, psychological screening, assessment, work intensity

Introduction

Chatbots have grown in popularity and acceptance in recent years. They are often used on websites to provide customer service, such as helping users navigate or answering Frequently Asked Questions. The advantage is that they are available 24/7 and can quickly answer recurring questions within a usually definable topic area, for instance, such as product features. As they have evolved, chatbots powered by artificial intelligence (AI) have taken on the role of virtual assistants, performing tasks such as scheduling appointments or initiating phone calls (Ramesh & Chawla, 2022). The rapid development of AI-powered chatbots has opened up a wide range of applications in psychology. For example, chatbots are used as supporting elements in therapeutic settings (e.g., Fitzpatrick, 2017). Chatbots are also an interesting area of research for industrial and organizational psychology, as they are increasingly accepted and integrated into the work context. They also offer a rich potential as a research tool, especially in the areas of diagnostics and intervention.

Talking to a chatbot is very similar to an interview, which is why initial studies have been conducted on the use of chatbots in the context of psychometric assessments. For example, chatbots have been shown to be useful assistants that can help complete a questionnaire (Söderström et al., 2021) or present existing questionnaires and ask which response category should be selected (e.g., Schick et al., 2022). Beyond simply assisting with questionnaire completion, an important question is whether chatbots could serve as an alternative method for collecting psychometric data (e.g., Fan et al., 2023).

In this article, we will look at the use of chatbots in the context of psychological assessment. First, we will describe how chatbots work as conversational agents and give a brief overview of the use of chatbots in psychological screening. Then, based on our own experience, we will describe a chatbot for assessing different facets of work intensity. Finally, we discuss our experience with the chatbot and provide an outlook on the use of chatbots in psychological screening.

Chatbots as conversational agents

A chatbot simulates a conversation by exchanging messages, either text-based or voice-based. Technically, a conversation is a sequence of recognising intentions and initiating corresponding responses or actions (e.g., Chakraborty et al., 2023). The logic of the conversation can be either rule-based or driven by AI. Rule-based chatbots rely on pre-defined scripts to recognise and respond to input. A very basic example would be the input “What is your name?”, which would result in the response “My name is James”. Because all scripts are pre-defined, rule-based chatbots are limited in their ability to recognise certain types of input, and their responses are limited to a pre-defined set of responses. Chatbots powered by AI can handle more complex conversations. In particular, intent recognition is more sophisticated and is handled by natural language understanding, which performs two tasks: intent classification and entity recognition.

Intent classification involves recognising what the user wants. For example, the user might ask what the weather will be like in Hamburg tomorrow. The user's intent is to know what the weather is like, and the chatbot concludes that it should report back a weather report. Entity recognition refers to instances such as names, places or time-based entities. In our example, the entities are a specific place (e.g., Hamburg) and a specific date (e.g., tomorrow). By combining the intent and the entities, the chatbot can infer that the user wants a weather forecast for Hamburg for the next day. Although this example sounds simple, the chatbot needs to recognise that the word Hamburg refers to a city. To do this, chatbots rely on dictionaries that list all possible cities. This can get complicated if words can be recognised as multiple entities. For example, a crane could be a bird, or a machine used to lift objects. Here the chatbot might infer from the context to identify the correct entity. If the word crane is used in the context of a discussion about eggs, the bird is more likely to be mentioned. If the word crane is mentioned with other materials, such as concrete or bricks, the machine is more likely.

Chatbots in psychological research

The development of chatbots is progressing at breakneck speed, and as their capabilities grow, they are becoming increasingly interesting for psychological research. This is not only about exploring the acceptance of chatbots, but also their use in healthcare, for example, where they could take over diagnosis and treatment (Chakraborty et al., 2023). Chatbots have also been tested in the context of cognitive behavioural therapy (Fitzpatrick, 2017) or in digital mental health interventions (Boucher et al., 2021). A comprehensive review by Casu et al. (2024) highlights how chatbots contributed to positive behaviour change and improved psychological well-being, demonstrating the growing role of these digital assistants in supporting mental health interventions. Complementing this research, Miner et al. (2019) explored the integration of AI-powered conversational systems in therapeutic contexts, examining how these technologies can augment traditional psychotherapy approaches. Together, these studies illustrate the evolving landscape of digital mental health tools and their potential to expand access to psychological support.

Chatbots in psychological assessment

There are already examples of chatbots being used in the context of psychological assessment. For example, Dos Santos de Lima et al. (2024) developed a chatbot that asked participants questions about physical and psychological symptoms of depression; the participants simply responded with "yes" or "no". Schick et al. (2022) presented a chatbot for mental health assessment that collects responses to closed-ended questions on a Likert scale from 1 to 5. Caballer et al. (2022) investigated the quantitative equivalence of a paper-and-pencil questionnaire and a voice-based conversational assistant. In particular, their chatbot administered the same questions from a paper-based questionnaire to assess loneliness and the results supported the feasibility of this approach. As another example, Podina et al. (2023) demonstrated that a chatbot system could accurately detect the presence or absence of clinical symptoms by analysing Likert scale responses through natural language processing (see also Boian et al., 2024). This approach has been used in various contexts, such as a mobile banking questionnaire (Celino & Re Calegari, 2020) or internet usage behaviour (Kim et al., 2019). However, Zarouali et al. (2024) found no evidence that chatbots could be a better survey research tool when compared to traditional web surveys.

The chatbots described so far have been used as delivery agents rather than conversational agents; they have only delivered closed-ended questions from traditional questionnaires and collected responses on Likert scales. However, simply using chatbots as delivery agents does not exploit their potential in psychological assessment. On the contrary, the use of open-ended questions, where the answer is interpreted by the chatbots, would be an innovative way to bridge the gap between quantitative and qualitative approaches to psychological assessment. To the best of our knowledge, Fan et al. (2023) is the only study that has used open-ended questions as an alternative approach to assess psychological constructs and comprehensively validated the assessment of the chatbot. In the following, we present a chatbot-based conversational survey designed to assess work intensity primarily through open-ended questions.

Chatbot on work intensity

We argue that chatbot technology has the potential to do more than just deliver questionnaire items. With this in mind, we have developed a chatbot designed to assess work intensity primarily through open-ended questions (Soucek & Voss, 2025). In a structured conversation, the chatbot asks participants to provide examples from their work experience, which are then used to assess seven different facets of work intensity. The assessment is based on natural language processing and intent recognition. To validate this approach, the ratings of the chatbot were compared with those of a validated questionnaire measuring the same facets of work intensity (Soucek & Voss, 2024).

Work intensity

The continuous advancement of information and communication technologies has changed and influenced the way we communicate and collaborate (Korunka & Kubicek, 2017). Work environments have evolved, offering greater flexibility in terms of both work location and working hours. In addition, as a result of the Covid-19 pandemic, it has become common for many employees to work from home and in the evenings. These new ways of working can contribute to increased work intensity (e.g., Soucek et al., 2024). Work intensity is often described as “working hard” including working for long hours at high levels of effort (Burke et al., 2010). However, the notion of “work at high speed and to tight deadlines might not fully capture the complexity of work intensity in its many job-specific forms” (Piasna, 2017, p. 171).

In an effort to gain a more comprehensive understanding of work intensity, Soucek and Voss (2024) conducted an extensive literature review and organised workshops with employees, employee representatives, and management (such as work councils and occupational health management). Based on this research, they developed a 21-item questionnaire that assesses seven different facets of work intensity. Specifically, work intensity is associated with a high number of tasks that occur in a given period of time (amount), that need to be done in parallel (concurrency), or that occur in temporary clusters (work peaks). Similarly, a high degree of coordination with colleagues is another characteristic of work intensity (interdependence). Other facets relate to a high degree of interruptions (interruptions), ambiguous tasks (lack of clarity) and availability for professional matters outside regular working hours (extended availability). This measure of work intensity has been shown to be relevant to emotional exhaustion and work engagement (Soucek et al., 2024).

Chatbot for the assessment of work intensity

As work intensity is a relevant and understandable topic for employees, it was chosen as the subject of a chatbot development. The chatbot was programmed using Rasa. Rasa is an open-source framework for building text- and speech-based applications (https://rasa.com/). Rasa provides a modular set of natural language understanding and dialogue management tools that enable the creation of advanced conversational AI systems. Another freely available option is the Python library ChatterBot (https://github.com/gunthercox/ChatterBot).

The conversation with the chatbot started when the user greeted it with a simple “hello”. The chatbot then responded with a friendly greeting and introduced the purpose of the conversation, namely work intensity. To demonstrate its ability to respond to user input, the chatbot asked how the user was feeling. Based on the user’s answer, it responded with either a supportive (“Wonderful!”) or encouraging message (“You'll be fine!”). After this initial exchange and demonstration of responsiveness, the chatbot moved on to the main discussion about work intensity.

The main part of the interview was divided into seven segments, each representing one of the seven aspects of work intensity (Soucek & Voss, 2024). Figure 1 shows the schematic sequence of the interview and the scoring of the first two facets. To avoid primacy or recency effects, the order of these segments was randomised for each participant. Each segment followed a consistent structure, beginning with an initial question taken from the questionnaire developed by Soucek and Voss (2024).

For each facet, we selected three questions from the Soucek and Voss (2024) questionnaire and rephrased them to allow a simple yes or no answer (e.g., interdependence: “Does working on your tasks require a lot of coordination with other people?”). For the initial question, one of these three questions was chosen at random. The chatbot used natural language understanding to interpret the answer as either agreement or disagreement. That is, the chatbot recognised positive responses not only when users explicitly said “yes”, but also when they used similar expressions such as “sure” or “absolutely”; the same was true for responses indicating a refusal. If the chatbot couldn’t clearly identify the answer, it randomly selected another question on the same facet as a fallback. If the chatbot detected a denial of the initial question, it proceeded to the next facet, starting with the initial question for that new facet.

If the chatbot detected agreement with the first question, it asked the participant twice to provide an example of a situation related to that question (e.g., “Can you think of a brief example of interdependence in your work? Please describe such a situation”). These follow-up questions were phrased differently each time to keep the conversation interesting and engaging. The chatbot used natural language processing to classify participants’ responses to these open-ended questions, ensuring that they aligned with the respective facet of work intensity.

As an initial question was asked for each facet, the chatbot was able to assess all seven facets of work intensity. If the initial questions were answered in the affirmative, the topic was explored in greater depth to assess the extent of work intensity. If suitable examples could be provided, this was taken as an indication of a high level of work intensity for that facet. If no suitable examples were provided, this could indicate that the work intensity for the facet was not too high.

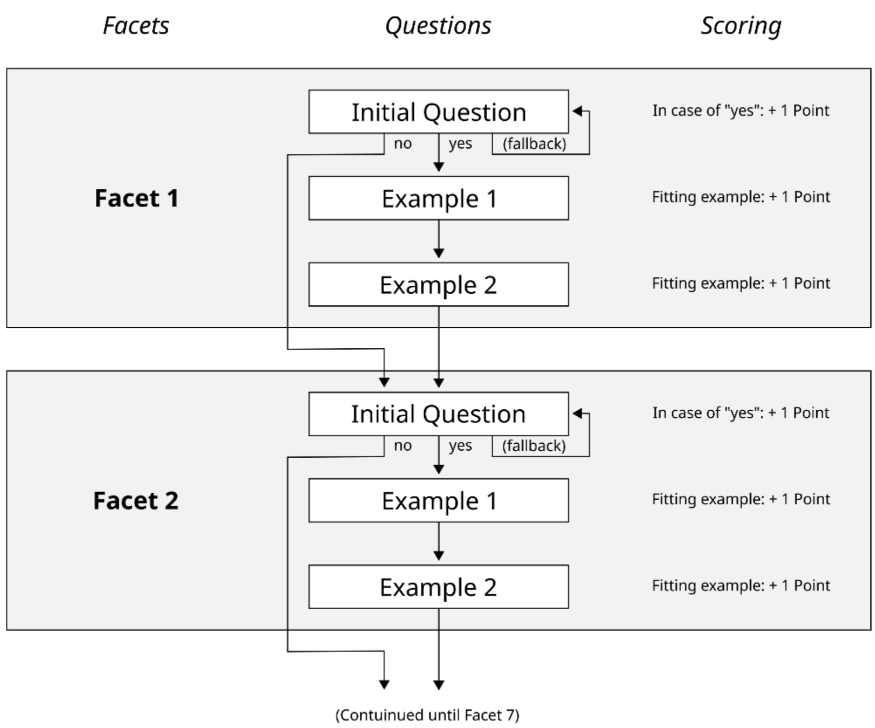

Figure 1. Schematic sequence of conversation and scoring

Validity of chatbot ratings

We conducted an evaluation study in which participants first interacted with the chatbot about work intensity and then completed a traditional questionnaire on the same topic (Soucek & Voss, 2025). In two separate studies, the chatbot's ratings showed the strongest correlations with the relevant facets as measured by the questionnaire, demonstrating convergent validity. In contrast, the correlations with unrelated facets were weaker, supporting discriminant validity. The second study with employees further confirmed the construct validity of the chatbot ratings, as they were significantly correlated with established measures of work intensity. Finally, criterion-related validity was assessed by examining the relationship between the chatbot ratings and outcomes such as emotional exhaustion and psychological detachment.

Interestingly, the correlations between the facets were weaker for the chatbot ratings compared to the questionnaire scores. As the questionnaire relied on self-reported data, the higher correlations may reflect a common method bias (Podsakoff et al., 2003). In contrast, the chatbot ratings were based on open-ended questions and an algorithm that was not obvious to participants, which is likely to minimise the impact of common method bias and socially desirable responses. Furthermore, asking participants for specific examples to illustrate each facet acts as an implicit form of validation. For example, if a participant affirms the initial question but doesn't provide a specific example, this could indicate that the facet is not strongly represented in their day-to-day work. While it is possible that the participant misunderstood the question, this would also prevent the facet from being overestimated. In addition, the chatbot's algorithm could explain the lower intercorrelations between the facets: it compares responses to a list of words to calculate the likelihood of a match with each facet. Interestingly, the chatbot assesses the likelihood for all facets at once, and a response is only assigned to a particular facet if it is sufficiently different from the others. Overall, the chatbot ratings seem to provide a more detailed and nuanced assessment of the seven facets of work intensity than the questionnaire (Soucek & Voss, 2025).

Experiences and perspectives

Experiences with the chatbot

We started developing the chatbot out of interest and were surprised at how well it worked. All in all, we had an initial version, which we improved after a first study. With the two subsequent studies we were able to confirm the factor structure and validate the ratings of the chatbot (Soucek & Voss, 2025). The combination of pre-defined intentions (i.e., the seven facets of work intensity) and qualitative analysis of the responses to the open-ended questions ensures that all seven facets are covered in the conversation. However, this strict structure of the interview may miss potentially relevant aspects that are not covered by the seven facets. By systematically analysing the data collected, new facets could be identified and incorporated into the model, improving the chatbot in the long term. In the future, this feedback could even lead to the adaptation of existing questionnaires.

We designed and programmed the chatbot's algorithm for assessing the seven facets of work intensity ourselves. This may be an initial hurdle, but there have been no comparable projects to date (the first version of the chatbot was developed in 2020). Meanwhile, technology has advanced, and new tools are being developed to simplify the programming of chatbots. We also had the advantage of having developed a questionnaire on the seven facets of work intensity (Soucek & Voss, 2024), so we were very familiar with the topic. Another advantage was certainly that the topic of work intensity is quite clear and understandable, although there are also subtle nuances between the individual facets. Psychological constructs, which are more abstract, may be harder to measure with a chatbot. In this case, other algorithms or big data analysis techniques may be more useful. Nevertheless, the development of a second chatbot to assess four facets of resilient behaviour using the same algorithm described above was quite promising (Soucek et al., 2015).

The structure of the conversation with the chatbot is based on semi-structured interviews, in which a quantitative opening is complemented by a qualitative approach. By asking participants to give examples from their everyday lives, they are encouraged to go deeper into the topic and may address aspects that would not be covered by a questionnaire. However, the scoring of responses and the detection of intentions is based on the comparison of probabilities, following a more quantitative approach. From this perspective, chatbot assessments could bridge the gap between qualitative and quantitative research.

Enhancing surveys with chatbots

A chatbot is a simple tool to invite employees to participate in surveys on work-related topics. Asking participants to give examples from their daily work could lead to a deeper engagement with the topic and thus contribute to an exchange of interpretations. From this point of view, the chatbot is not just a diagnostic tool, but already contains aspects of intervention.

Chatbots can be a more inviting and engaging assessment method than traditional questionnaires. Administering the survey through a human-like social interaction may encourage more information sharing, reduce survey fatigue and improve the quality of responses (Jin et al., 2024). In this regard, Xiao et al. (2020) found that participants spent more time, wrote longer responses, and disclosed more information. Thus, interacting with a chatbot could increase participants' engagement. Respondents who engaged with the chatbot-led survey provided responses that were more relevant, specific and clear, indicating higher response quality (Xiao et al., 2020).

Another aspect is the user's perception of the chatbot. One way to increase participants' engagement in communicating with chatbots could be through the design of the chatbots. For example, Rhim et al. (2022) incorporated the humanisation techniques of self-introduction by the chatbot, addressing the interlocutor by name, using adaptive response speed, and echoing respondents' responses. Rhim and colleagues (2022) compared this humanised chatbot with a basic chatbot without these techniques. The results showed that respondents' perceptions were more positive and they reported higher levels of satisfaction. In addition, respondents spent more time with the chatbot and self-disclosure was higher. However, social desirability bias was higher compared to the baseline chatbot (Rhim et al., 2022). While some degree of humanisation may increase acceptance, over-humanisation carries the risk of the *uncanny valley hypothesis* (Mara et al., 2022). This hypothesis suggests that artificial characters that are almost, but not quite, human-like can cause a deep sense of discomfort or unease. Accordingly, chatbots should remain recognisable as such and not be overly humanised.

Ethical issues related to chatbots

While the use of chatbots holds great promise, it also presents challenges and requires well-defined guidelines. For example, concerns have been raised about limitations and ethical implications, particularly in relation to the concept of therapeutic misunderstanding (Khawaja & Bélisle-Pipon, 2023).

A practical example of ethical issues arose from our own experience in developing the chatbot. As the conversation was in German, we used German dictionaries for intent recognition. We also experimented with entity recognition. In particular, we wanted the chatbot to recognise first names so that we could address the user by name. This also works well with traditional German first names such as "Hildegard". However, the chatbot had problems recognising names that were not of German origin, such as "Amanda". So even small things like this need to be considered during development, otherwise the chatbot runs the risk of being biased by systematically excluding certain groups.

Based on our experience, we strongly recommend a deliberate and controlled approach to the development of chatbots for use with psychological questions. Diaz-Asper et al. (2024) developed a framework for ethical challenges in the use of speech technologies in psychometric research. One of the main challenges is to translate broad ethical principles into concrete, context-specific guidelines. For example, respondents who require further investigation or intervention should not be left to chatbots. Those providing chatbots in practice should have strategies in place to extract relevant action points from conversations with chatbots, rather than completely outsourcing the discussion of problems to chatbots.

Conclusion

The use of chatbots in psychometric research has great potential, especially through the combination of quantitative and qualitative elements. However, chatbots are not intended to replace questionnaires in psychometric assessment, but could compensate for some of their disadvantages in certain scenarios. From a practical perspective, they offer a low-threshold and intuitive approach to topics such as work intensity and are therefore a promising tool for initial screening.

References

Boian, R., Bucur, A.-M., Todea, D., Luca, A., Rebedea, T., & Podina, I. R. (2024). A conversational agent framework for mental health screening: Design, implementation, and usability. *Behaviour & Information Technology*, 1–15. https://doi.org/10.1080/0144929x.2024.2332934

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, *18*(sup1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Burke, R. J., Singh, P., & Fiksenbaum, L. (2010). Work intensity: Potential antecedents and consequences. *Personnel Review*, *39*(3), 347–360. https://doi.org/10.1108/00483481011030539

Caballer, A., Belmonte, O., Castillo, A., Gasco, A., Sansano, E., & Montoliu, R. (2022). Equivalence of chatbot and paper-and-pencil versions of the De Jong Gierveld loneliness scale. *Current Psychology*, *41*(9), 6225–6232. https://doi.org/10.1007/s12144-020-01117-0

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI chatbots for mental health: A scoping review of effectiveness, feasibility, and applications. *Applied Sciences*, *14*(13), 5889. https://doi.org/10.3390/app14135889

Celino, I., & Re Calegari, G. (2020). Submitting surveys via a conversational interface: An evaluation of user acceptance and approach effectiveness. *International Journal of Human-Computer Studies*, *139*, 102410. https://doi.org/10.1016/j.ijhcs.2020.102410

Chakraborty, C., Pal, S., Bhattacharya, M., Dash, S., & Lee, S.-S. (2023). Overview of chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1237704

Diaz-Asper, C., Hauglid, M. K., Chandler, C., Cohen, A. S., Foltz, P. W., & Elvevåg, B. (2024). A framework for language technologies in behavioral research and clinical applications: Ethical challenges, implications, and solutions. *American Psychologist*, *79*(1), 79–91. https://doi.org/10.1037/amp0001195

Dos Santos de Lima, E., Scortegagna, S. A., & Bertoletti de Marchi, A. C. (2024). Psico Bot: a robot for psychological assessment of children and adolescents. *Ciencias Psicológicas*. https://doi.org/10.22235/cp.v18i2.3428

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology*, *108*(8), 1277–1299. https://doi.org/10.1037/apl0001082

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Jin, Y., Chen, L., Zhao, X., & Cai, W. (2024). The way you assess matters: User interaction design of survey chatbots for mental health. *International Journal of Human-Computer Studies*, *189*, 103290. https://doi.org/10.1016/j.ijhcs.2024.103290

Khawaja, Z., & Bélisle-Pipon, J.-C. (2023). Your robot therapist is not your therapist: Understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1278186

Kim, S., Lee, J., & Gweon, G. (2019). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300316

Korunka, C., & Kubicek, B. (2017). *Job demands in a changing world of work*. Springer. https://doi.org/10.1007/978-3-319-54678-0

Mara, M., Appel, M., & Gnambs, T. (2022). Human-like robots and the uncanny valley: A meta-analysis of user responses based on the godspeed scales. *Zeitschrift für Psychologie*, *230*(1), 33–46. https://doi.org/10.1027/2151-2604/a000486

Miner, A. S., Shah, N., Bullock, K. D., Arnow, B. A., Bailenson, J., & Hancock, J. (2019). Key considerations for incorporating conversational AI in psychotherapy. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00746

Piasna, A. (2017). Scheduled to work hard: The relationship between non-standard working hours and work intensity among European workers (2005-2015). *Human Resource Management Journal*, *28*(1), 167–181. https://doi.org/10.1111/1748-8583.12171

Podina, I. R., Bucur, A.-M., Fodor, L., & Boian, R. (2023). Screening for common mental health disorders: A psychometric evaluation of a chatbot system. *Behaviour & Information Technology*, 1–10. https://doi.org/10.1080/0144929x.2023.2275164

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Ramesh, A., & Chawla, V. (2022). Chatbots in marketing: A literature review using morphological and co-occurrence analyses. *Journal of Interactive Marketing*, *57*(3), 472–496. https://doi.org/10.1177/10949968221095549

Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior*, *126*(107034), 107034. https://doi.org/10.1016/j.chb.2021.107034

Schick, A., Feine, J., Morana, S., Maedche, A., & Reininghaus, U. (2022). Validity of chatbot use for mental health assessment: Experimental study. *JMIR mHealth and uHealth*, *10*(10), e28082. https://doi.org/10.2196/28082

Söderström, A., Shatte, A., & Fuller-Tyszkiewicz, M. (2021). Can intelligent agents improve data quality in online questionnaires? A pilot study. *Behavior Research Methods*, *53*(5), 2238–2251. https://doi.org/10.3758/s13428-021-01574-w

Soucek, R., Pauls, N., Ziegler, M., & Schlett, C. (2015). Entwicklung eines Fragebogens zur Erfassung resilienten Verhaltens bei der Arbeit [Development of a questionnaire for the assessment of resilient behavior in the workplace]. *Wirtschaftspsychologie*, *17*(4), 13–22.

Soucek, R., Voss, A. S., Drexler, H., & Moser, K. (2024). New ways of working and psychological well-being: Work intensity as a target variable of job resources. *International Journal of Workplace Health Management*, *17*(4), 353–366. https://doi.org/10.1108/IJWHM-01-2024-0010

Soucek, R., & Voss, A. S. (2024). Rethinking the assessment of work intensity: Development and validation of a verbal questionnaire and pictorial scales. *Zeitschrift für Arbeits- und Organisationspsychologie*, *68*(4), 188–199. https://doi.org/10.1026/0932-4089/a000433

Soucek, R., & Voss, A. S. (2025). *Let's talk about your work: Validation of a chatbot for the assessment of work intensity*. Manuscript submitted for publication.

Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction*, *27*(3), 1–37. https://doi.org/10.1145/3381804

Zarouali, B., Araujo, T., Ohme, J., & de Vreese, C. (2024). Comparing chatbots and online surveys for (longitudinal) data collection: An investigation of response characteristics, data quality, and user evaluation. *Communication Methods and Measures*, *18*(1), 72–91. https://doi.org/10.1080/19312458.2022.2156489