Measuring the performance of human-machine symbiosis

Emergent Research Forum (ERF)

Patrick Derckx
University of Cologne
derckx@wim.uni-koeln.de

Franziska Kraken
University of Cologne
kraken@wim.uni-koeln.de

Detlef Schoder
University of Cologne
schoder@wim.uni-koeln.de

Abstract

In today's economy, machines are evolving and becoming increasingly important for supporting human's daily life and work. Recently, scholars even discovered a symbiotic relationship between humans and machines, learning from each other to improve their joint processes and outputs. However, to our knowledge, this effect has not yet been rigorously measured. With this research in progress, we directly address this gap and propose to create a method consisting of evaluation criteria as well as an instrument for gathering data to measure human-machine symbiosis by interrogating users.

Keywords

Human-Machine Symbiosis, Design Research, Collaboration, Artificial Intelligence, Evaluation.

Introduction

The world has arrived in the second age of machines (Brynjolfsson and McAfee 2014). From simple steam machines in the 19th century to today's sophisticated, AI-driven machines, we observe a tremendous enhancement. Once, industrialization brought our workers from the fields to the factories and created an infrastructure that made our modern (Western) society and the associated high standard of living possible. Luxury goods that for generations were available to only a few are now affordable for the average citizen. And this has evolved mainly through an increase in efficiency, which went hand in hand with the introduction of very simple industrial machines. These industrial machines have evolved greatly since then. They have become more precise, faster, larger and more reliable. Nevertheless, for a long time they remained nothing more than the human's "henchman", carrying out the (highly repetitive) routine tasks assigned to them. Today, the world is once again facing major changes. Through the latest developments in the field of Artificial Intelligence (AI), machines are better able to understand situations and react accordingly. They are better able to incorporate creativity, intuition and human sense into their work and solve tasks together with humans. Considerations on such human-machine symbiosis (HMS) go back to the 1960s (Licklider 1960). However, through the latest developments, it is possible not only to describe these systems but also to build them. The literature has already presented the first symbiotic systems, for instance in the field of logistics (Döppner et al. 2018)or the Startup Success Prediction (Dellermann et al. 2017). But evaluation of these systems, is limited strictly to each individual case and its performance indicators. Current research does not investigate the collaboration or the relationship between the actors, or only does so very limited. And this despite the fact that the first studies prove and describe co-evolution, which is essential for a symbiosis (Döppner et al. 2019). In HMS, there are more far-reaching effects that go beyond simple performance indicators. It is therefore important to evaluate HMS inside the socio-technical system (Geels 2004). Such a contribution is missing; hence the following general research question arises: *How does HMS perform?* Since this research question is very abstract, we are narrowing the scope of this research project and focusing on developing a strategy for the evaluation of HMS and testing it by one instantiation. As e.g. Prat et al. (2015) have shown that evaluation strategies for Design Science Research

(DSR) artefacts already exist, this leads to the narrower research question that forms the basis of the proposed research project where it is proposed that such a strategy be developed for HMS: ***R1****: Which evaluation strategy is appropriate to measure HMS performance?* This research in progress addresses exactly the research gap described earlier. We will create a strategy containing a technique as well as the individual criteria derived from the literature on collaborative systems and DSR to evaluate HMS. Moreover, dealing with the evaluation of artefacts itself is an important contribution in Information Systems (IS) science, as this is a very under-researched field.

Background

In this section, we give an introduction to the relevant literature strands for this research project. We start by explaining what we understand by HMS. Then we describe what evaluation strategies mean for DSR artefacts. Finally, we briefly present another research project in which a DSS was developed according to the concepts of HMS.

Human-Machine Symbiosis

From a morphological perspective, the term HMS does not simply correspond to a translation of Licklider's Man-Computer Symbiosis, but rather consists of three parts: (1) Human; the simple use of the term 'man' would be politically incorrect (Griffith 2007). (2) Machine; a machine is "an apparatus, consisting of a number of interrelated parts, constructed to perform a task" ("Machine, n." 2019). In this way, every technological developments such as software, computers, robots, smartphones and virtual-reality glasses can be considered as partners for the human (Maier et al. 2018). (3) Symbiosis; the botanist de Bary introduced the term in 1879 to describe any type of coexistence of different organisms (1879). According to the above explanations, symbiosis is described as the coexistence of actors of different kinds for mutual benefit. In our previous work, we presented a comprehensive conceptualization of HMS (Gerber et al. 2020). For the proposed research, the fundamentals shortly described below are of particular relevance.

- **Objectives**: The main goal of HMS is coexistence for all actors with mutual benefit. All actors are seen together as part of an effective system (Jacucci et al. 2014). In this system, all actors pursue a common goal (Döppner et al. 2018). To reach this goal, the limits of human trade are overcome by an optimal division of labour (Haasbroek 1993). The main aim is to save time in the execution of tasks. All actors are largely equal partners (Sandini et al. 2018).
- **Requirements**: The number of actors in a HMS is not specified in great detail in the literature but is usually implicitly referred to as a 1:1 relationship (Sanchez et al. 2009). The role of the machine in this relationship must be equal to that of the human being (Jarrahi 2018). Control must not emanate from any side (Sandini et al. 2018). The actors must cooperate effectively as partners (Jarrahi 2018). The machine must have a natural character for humans to perceive it as a partner. The system must be dynamic, i.e. the actors must be able to adapt to new roles and tasks (Sanchez et al. 2009). However, the task and division of labour must be clearly defined at all times (Haasbroek 1993).
- **Limits**: In the literature, limits for HMS are described; these are mainly obstacles that lead to objectives not being achieved or requirements not being met (e.g. the lack of a clear objective leads to a common goal not being achived). For a real HMS, it is existential to overcome these limits. Special mention is made of trust in the machine and data security concerns.

Evaluation of Artefacts in IS Research

In this paper, we deal with the evaluation of artefacts. Hevner's DSR Framework (2004) enabled relevant and rigorous design-oriented research for in IS research. Creating artefacts implies the development of theories and concepts above the instantiation. In order to test whether the artefact meets the requirements of the corresponding theory or concepts, hypotheses are derived and tested in an evaluation (Prat et al. 2015). If the artefact follows an established theory or concept, the DSR scientist can prove the usefulness of his artefact (Hevner et al. 2004). However, there are also DSR projects that do not design artefacts on the basis of an established theory in order to use them in a suitable application case. Rather, theories and concepts from other areas are adapted and tested in the new application context, or completely new ones are developed. The evaluation of DSR artefacts serves not only to check against the a theory or a concept,

but also to evaluate the theories themselves. Prat et al. (2015) subdivide the components of the evaluation into evaluation criteria (the 'what' of evaluation) and techniques (the 'how' of evaluation).

Evaluation criteria: Prat et. al describes five criteria categories (2015): The *goal* category includes criteria that measure the usefulness of the artifact. In addition, the achievement of objectives is evaluated and whether the development of the artifact is technically, operationally and economically possible is assessed. It also examines whether the findings can be generalized. The effects of the artifact on the context are evaluated in the *environment* category. Questions about usability by the user, embedding in the organization and the appropriateness between technology and corporate culture are answered. *Structure* means the simplicity and comprehensibility of the artifact, as well as the degree of adaptation of the artifact to the underlying theories and concepts from research. In the *activity* category, the process of use is evaluated. It checks whether the process is simple, consistent and self-explanatory through the use of the artifact. In addition, the trustworthiness of the artifact is evaluated. In the last category, criteria are summarized that concern the *evolution* of the artifact. Thus, the robustness of the artifact, its scalability and adaptability are evaluated. The learning ability of the artifact is also evaluated.

Evaluation methods (EM): EM, describe the way in which evaluation is carried out. In the DSR literature, this is still little investigated. Nevertheless, some techniques are presented by Hevner et al. and Peffers et al. (2004; 2007) e.g. case study, field study, static analysis or simulations (Prat et al. 2015). It is not necessary to use one technique for the entire evaluation, but rather the techniques can be mixed and the choice of appropriate method made depending on the criteria. In particular, little attention is paid to the link between, and the combination of, methods and criteria – what we name evaluation strategy. Venable presents a framework which brings together two dimensions of evaluation and tries to put the 'how' and 'what' of evaluation into context. They define the dimensions 'naturalistic' vs. 'artificial' and 'ex ante' vs. 'ex post', which lie opposite each other on a continuous scale (2012). *Pur naturalistic* preys on the fact that real users use real artefacts and thereby solve real problems (Venable et al., 2012). On the other hand, *pur artifical* means that the setting is constructed. The evaluation can take place continuously during the development of the artifact (*ex ante*) or downstream (*ex post*). Prat et al. go further and describe five dimensions of evaluation (2015). The evaluation technique can be observational, e.g. with case or field study or experimental. The last includes a controlled experiment or a simulation. The technique can also be analytical, descriptive or question based. The Form of Evaluation differs mainly according to whether it follows a qualitative or quantitative approach (Cleven et al. 2009), is logically argumentative or formally proven. Secondary participants are those persons with whom the evaluation is conducted who were not involved in the development. The Level of Evaluation dimension corresponds to the 'ex ante vs ex post' dimension also proposed by Veanble et al. (2012). The ex-ante evaluation considers an abstract artifact, whereas the ex post evaluation deals with an instanced artifact.

An Intelligent Decision Support System for the Air Cargo Industry

With a partner from the air cargo industry, we have been working for several years on the development of a decision support system (DSS). Our partner provides an outsourcing service for the management of containers and pallets used in air cargo (ULDs – Unit Load Devices). The company ensures the ULD supply for their customers while being responsible for a fleet of more than 100,000 ULDs of multiple airlines operating worldwide. ULD dispatchers continuously control the network and reallocate the ULD stocks. This problem is called the Empty ULD Repositioning (EUR) problem (Döppner et al. 2018). Before a DSS was introduced, decision making was done manually. With the constructed artifact, parts of this decision-making process can be done by the machine. The project is set up as a multi-year DSR project, currently in its fourth iteration loop. First, we focused on the development of a DSS to prove that the complex decision making of EUR can be supported. Afterwards, we focus on the division of labour between human and machine and the identification of a purposeful level of task automation. Thirdly, we focus on an emerging co-evolution that becomes apparent when ULD dispatchers interact with the artefact (Döppner et al. 2019).

Methodology

We will solve the research question by applying an inductive and deductive approach. In concrete terms, this means that we develop an abstract evaluation strategy for the HMS using DSR after Hevner and Peffers (2004; 2007) . Afterwards, we use a concrete instantiation of a DSR project to apply the evaluation strategy

using the project described in section 2 and deductively evaluate the artefact corresponding HMS. We start with the development of a suitable evaluation strategy applying the following steps:

- Selection of evaluation techniques: An evaluation strategy always consists of the 'how' and the 'what' of the evaluation. The evaluation techniques, the 'how', are given by the concrete instantiation. We therefore do not make any specifications here.
- Selection of evaluation criteria: We use the general established criteria from the taxonomy of Prat et al. (2015) as a starting point for the selection of suitable criteria. To find out which criteria are relevant for the HMS, we compare the taxonomy with the concepts of the HMS (Gerber et al. 2020) and enrich them with the few HMS instantiations from the literature.
- Development of measures: In order to evaluate the HMS, we need to show which proxy is used to evaluate the individual criteria. To do this, we check which data sources are suitable. As data sources, we want to use data from project documentation, metrics from system usage and user experiences from interviews.
- Instrument development: As an additional data source, we create a questionnaire to evaluate the criteria appropriately.

The previous will result in a robust evaluation strategy containing the method (how), a set of evaluation criteria (what) and an instrument to collect data. Afterwards the evaluation will be carried out using our example project. We will check if the data sources and the instrument offer the relevant data to measure the criteria. If problems arise, we iterate our evaluation strategy. At the end, we communicate our evaluation results to the community in addition to the strategy.

Current State of the Research Project

In section 2, we have shown that we have gained both a broad and deep understanding of the issues addressed in this project: HMS and DSR evaluation. With our conceptualization presented in section 2, we presented concepts of the HMS in detail and can therefore properly describe what we want to investigate. We have conducted a literature review of DSR evaluation to identify which evaluation strategies existing. The papers identified were mostly cited in section 2. We could not identify evaluation strategies for specific concepts (such as HMS). We confirmed that the evaluation of artefacts in DSR is, a scarcely considered in the IS literature. Currently we are deriving the criteria relevant to HMS and are developing a tool to quantify them. In doing so, we are strongly oriented towards Prat et. al (2015). For space reasons, we have excluded the resulting preliminary questionnaire and the preliminary criteria set from this work. With this article outlining our research in progress, we would like to share and discuss our idea and the research method with an international audience. Therefore, it was important to present our motivation and especially our preliminary work from the literature. We hope to incorporate our findings and discussions in a subsequent full paper on the evaluation strategy, the instrument and the criteria set, once they are completed and tested.

Conclusion

In this research in progress, we have outlined the possibilities for evaluating artefacts and in particular how evaluation can be applied to HMS. We have selected a robust set of evaluation criteria from the relevant literature and compared them with the concepts of HMS. Research in the field of HMS is still in its infancy and is very conceptually oriented. There are only a few artefacts that claim the concepts for themselves and most of the existing ones are only sparsely evaluated, if at all. For future research in the field of HMS, we offer a clearly understandable selection of criteria for the evaluation of HMS to make the evaluation easier, so that hopefully more researchers may perform an artefact evaluation. In addition, we follow the call of Prat et al. (2015), who remark that their taxonomy should not be regarded as complete. We extend their work not only according to criteria, especially for HMS, but also generally according to the area of cooperation, which has not yet been considered in the literature until now. Besides that, the evaluation of artefacts is given little attention in the DSR literature. We are making a further contribution to this area. By also using our current state of research, practitioners and other researchers can use our evaluation criteria to carry out their own evaluations. Using these possible future evaluations, auctioneers can even learn and assess whether the situation is suitable before the start of an implementation.

Nevertheless, our study is not free of limitations. In the introduction, we posed the general research question of how HMS performs. By means of the strategy we have designed, this can be described for each

specific artifact. By conducting the evaluation, we can both test our strategy and provide indications for answering the general question. But it can only be answered if other scientists take up our work, follow it up and evaluate further artefacts.

REFERENCES

Bary, A. de. 1879. *Die Erscheinung der Symbiose: Vortrag, Gehalten auf der Versammlung Deutscher Naturforscher und Aerzte zu Cassel*, Strassburg: Verlag von Karl J. Trübner.

Brynjolfsson, E., and McAfee, A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, (1st ed.), W. W. Norton & Company.

Cleven, A., Gubler, P., and Hüner, K. M. 2009. "Design Alternatives for the Evaluation of Design Science Research Artifacts," in *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, DESRIST '09, New York, NY, USA: ACM, 19:1–19:8.

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., and Leimeister, J. M. 2017. *Finding the Unicorn: Predicting Early Stage Startup Success through a Hybrid Intelligence Method*, presented at the International Conference on Information Systems (ICIS), Seoul, South Korea.

Döppner, D. A., Derckx, P., and Schoder, D. 2018. "An Intelligent Decision Support System for the Empty Unit Load Device Repositioning Problem in Air Cargo Industry," in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS 2018)*, Big Island, Hawaii, USA.

Döppner, D. A., Derckx, P., and Schoder, D. 2019. "Symbiotic Co-Evolution in Collaborative Human-Machine Decision Making: Exploration of a Multi-Year Design Science Research Project in the Air Cargo Industry," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Geels, F. W. 2004. "From Sectoral Systems of Innovation to Socio-Technical Systems: Insights about Dynamics and Change from Sociology and Institutional Theory," *Research Policy* (33:6), pp. 897–920.

Gerber, A., Derckx, P., Döppner, D. A., and Schoder, D. 2020. "Conceptualization of the Human-Machine Symbiosis – A Literature Review," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Maui, Hawaii, USA.

Griffith, D. 2007. "Neo-Symbiosis: A Conceptual Tool for System Design," in *Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS)*, IEEE, January.

Haasbroek, L. J. 1993. "Advanced Human-Computer Interfaces and Intent Support: A Survey and Perspective," in *Proceedings of IEEE International Conference on Systems Man and Cybernetics Conference (SMC)*, Le Touquet, France: IEEE, pp. 350–355.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Q.* (28:1), pp. 75–105.

Jacucci, G., Spagnolli, A., Freeman, J., and Gamberini, L. 2014. "Symbiotic Interaction: A Critical Definition and Comparison to Other Human-Computer Paradigms," in *Symbiotic Interaction 2014* (Vol. 8820), G. Jacucci, L. Gamberini, J. Freeman, and A. Spagnolli (eds.), Springer Intern. Publishing, pp. 3–20.

Jarrahi, M. H. 2018. "Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making," *Business Horizons* (61:4), pp. 577–586.

Licklider, J. C. R. 1960. "Man-Computer Symbiosis," *IRE Transactions on Human Factors in Electronics* (HFE-1:1), pp. 4–11.

"Machine, n." 2019. *OED Online*, Oxford University Press. (www.oed.com/view/Entry/111850).

Maier, M., Ebrahimzadeh, A., and Chowdhury, M. 2018. "The Tactile Internet: Automation or Augmentation of the Human?," *IEEE Access* (6), pp. 41607–41618.

Peffers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *J. Manage. Inf. Syst.* (24:3), pp. 45–77.

Prat, N., Comyn-Wattiau, I., and Akoka, J. 2015. "A Taxonomy of Evaluation Methods for Information Systems Artifacts," *Journal of Management Information Systems* (32:3), pp. 229–267.

Sanchez, J. C., Mahmoudi, B., DiGiovanna, J., and Principe, J. C. 2009. "Exploiting Co-Adaptation for the Design of Symbiotic Neuroprosthetic Assistants," *Neural Networks* (22:3), pp. 305–315.

Sandini, G., Mohan, V., Sciutti, A. M., and Morasso, P. 2018. "Social Cognition for Human-Robot Symbiosis-Challenges and Building Blocks," *Frontiers in Neurorobotics* (12:34), pp. 1–19.

Venable, J., Pries-Heje, J., and Baskerville, R. 2012. "A Comprehensive Framework for Evaluation in Design Science Research," in *Design Science Research in Information Systems. Advances in Theory and Practice*, Lecture Notes in Computer Science, K. Peffers, M. Rothenberger, and B. Kuechler (eds.), Springer Berlin Heidelberg, pp. 423–438.