Industrial Cognitive Ergonomics and Engineering Psychology, Vol. 35, 2022, 61–67
https://doi.org/10.54941/ahfe1001602

Trust in an Autonomous Agent for Predictive Maintenance: How Agent Transparency Could Impact Compliance

Simon Loïck, Rauffet Philippe, Guérin Clément, and Seguin Cédric

Lab-Sticc, Université Bretagne Sud, UMR 6285 56100, Lorient, France

ABSTRACT

Human-machine cooperation is more and more present in the industry. Machines will be sources of proposal by giving human propositions and advice. Humans will need to make a decision (comply, i.e., agree, or not) with those propositions. Compliance can be seen as an objective trust and experiments results unclear about the role of risk in this compliance. We wanted to understand how transparency on reliability, risk or those two in addition will impact this compliance with machine propositions. With the use of an AI for predictive maintenance, we asked participants to make a decision about a proposition of replanification. Preliminary results show that transparency on risk and total transparency are linked with less compliance with the AI. We can see that risk transparency has more effect on creating an appropriate trust than reliability transparency. As we see, and in agreement with recent studies, there is a need to understand at a finer level the impact of transparency on human-machines interaction.

Keywords: Transparency, Human-machine interaction, Compliance, Trust

INTRODUCTION

In the context of Industry 4.0, human operators will increasingly cooperate with intelligent systems, considered as teammates in the joint activity (Romero et al, 2016). This human-autonomy teaming is particularly prevalent in the activity of predictive maintenance, where the system advises the operator to advance or postpone some operations on the machines according to the projection of their future state. Like in human-human cooperation, the effectiveness of cooperation with those autonomous agents especially depends on the notion of trust (Hoffman et al., 2013). The challenge is to calibrate an appropriate level of trust and avoid misuse, disuse, or abuse of the recommending system (Parasuraman & Riley, 1997). Compliance (i.e., positive response of the operator on a proposition or an advice, from an autonomous agent) can be interpreted as an objective measure of trust as the operator relies on the proposition from the autonomous agent (Chen, Mishler & Hu, 2021; Wang, Pynadath & Hill, 2016).

Recent studies propose a way to calibrate the trust by using the transparency concept (de Visser et al., 2020). Transparency has been defined as an information during a human-machine interaction that is easy to use with the intent to promote the comprehension, the shared awareness, the intent, the

role, the interaction, the performance, the future plans and the reasoning process (Roundtree, Goodrich & Adams, 2019; Chen & al, 2018; Lyons, 2013). Therefore, this research will focus on two aspects of the transparency concept: 1) Reliability transparency is information on likelihood of success/failure of the autonomous agent. It permits to have access on the reliability of the autonomous agent. This transparency is the probability of the autonomous agent to be right about its proposition or to succeed its task. For example, the autonomous agent will communicate that there is 90% chance of him being correct about its proposition. 2) Risk transparency is information on the projection of future outcomes. It permits to have access to the consequences that the autonomous agent perceives about its proposition or its task. For example, the autonomous agent will communicate that there is a risk to damage an equipment with its proposition. These two pieces of information are part of the Situation Awareness based-agent Transparency proposed by Chen and al. (2018).

Comprehension of what compliance is based on is still needed. Based on the model proposed by Chancey et al. (2017) this compliance can be mitigated by the risk perception (i.e., how humans interpret the consequence of the signal emitted by a machine). Nonetheless, we did not find any studies that tried to explore that question to highlight the role of risk in compliance.

The objective of this research is to understand the effect of the autonomous agent transparency on human compliance after a proposition from an autonomous agent (here an AI for predictive maintenance) for a more or less complex situation. We also wanted to understand at a finer level the impact of different transparency. We presented here our hypotheses for this research:

Hypothesis: Risk transparency will decrease compliance

Hypothesis: Reliability transparency will increase compliance

Hypothesis: Full transparency will decrease compliance

METHODS

We recruited 39 participants (mean age : 22 years, std = 2,7) from an engineering formation (mechatronics and industrial engineering) and they were compensated with a 10 euros gift card. After a brief formation of maintenance in maritime context, their role and their objectives for the experiment, participants had to familiarize themselves with the interface (Fig. 1).

Personality measures were assessed as control factors (Affinity to technology (Franke et al., 2019), Propensity to trust technology (Jessup et al., 2018) and Risk Propensity (Zhang et al., 2019)). After that, they were asked to complete height decision situations. Participants had to accept or deny a proposition, from a predictive maintenance algorithm, of advancing or postponing a CMMS[1] maintenance (Fig. 2).

Repeated measures of trust (based on Mayer, Davis & Schoorman, 1995 and Lyons & Guznov, 2019) and risk perception (based on Wilson, Zwickle and Walpole, 2018) were assessed after each situation. During this experiment, agent transparency level was manipulated by displaying information

[1] Computerized Maintenance Management System.

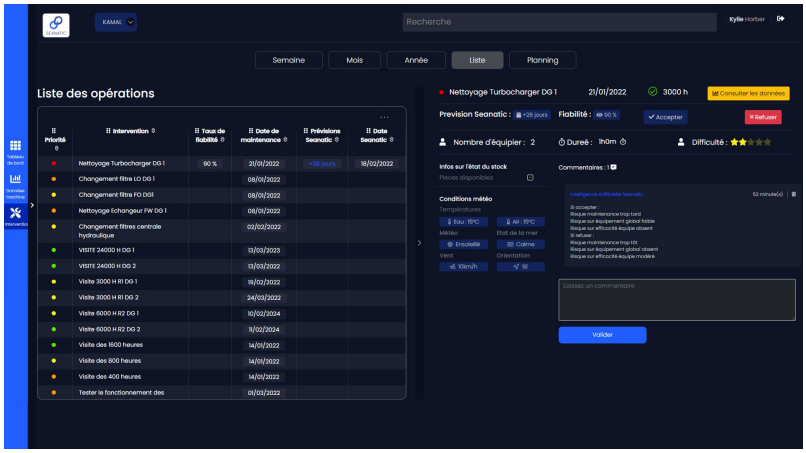

Figure 1: Seanatic interface used in the experiment.

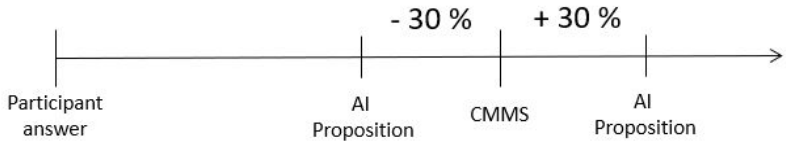

Figure 2: Timeline of AI proposition.

related to agent reliability, fixed at 90% (reliability transparency), and to situation outcomes (risk transparency), separately or in combination (full transparency). Risk was presented as a text based on three dimensions (risk on maintenance, impact on equipment and impact on team schedule). Participants had also access to the data used by the AI to make its proposition. This agent transparency was mixed with situation complexity (high or low) and the type of proposition (advancing or postponing the maintenance interventions). Complexity of the situation was defined by the criticality of the piece equipment on the global equipment (for example: the change of an oil filter has more impact for the global health of the equipment, and it needs to be done more frequently compared to the impact and frequency of cleaning the turbocharger). For the type of proposition, we decided to fix the proposition at -30% or +30% of the CMMS (for example, if due date maintenance was every 100 hours, the proposition was at 70 hours or at 130 hours). We coded the compliance as when the participant accepts the proposition of the AI. Personality measure and repeated measures (i.e., trust, risk perception and workload) will not be treated in this paper.

RESULTS

We will present here the primary results as the approfunding results are still an ongoing work. Contingency tables, and associated graphics have been used to see if there was a difference between compliance in the different experimental situations. We can observe that when the AI is transparent about the

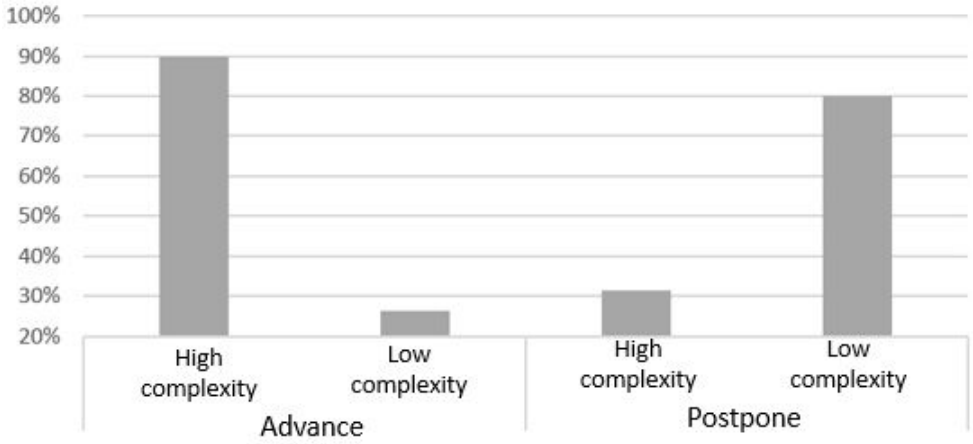

Figure 3: Compliance with AI proposition for risk transparency, high/low complexity and advancing (-30) or postponing (30) the operation.

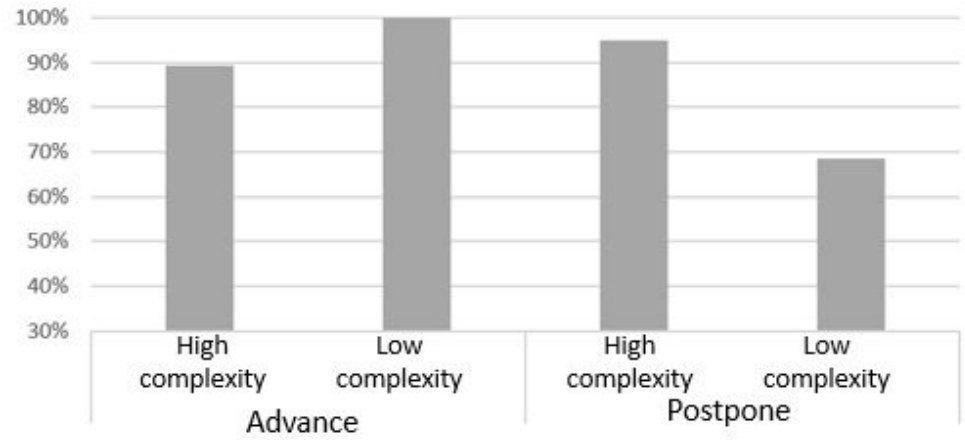

Figure 4: Compliance with AI proposition for reliability transparency, high/low complexity and advancing (-30) or postponing (30) the operation.

risk, participants were less compliant in the low complexity scenario of advancing propositions (26% against 80%) and in the high complexity scenario of postponing operation (31% against 80%) (Fig 3.)

We can observe that when the AI was transparent about the reliability, participants were very compliant with the proposition, no matter the complexity of the scenario nor the type of proposition (90% for high complexity and 100% for low complexity for advancing; 95% in high complexity and 68% in low complexity for postpone) (Fig. 4).

We can observe that when the AI was totally transparent, participants were less compliant in the low complexity scenario of advancing propositions (71% against 97%) and in the high complexity scenario of postponing operation (48% against 74%) (Fig. 5).

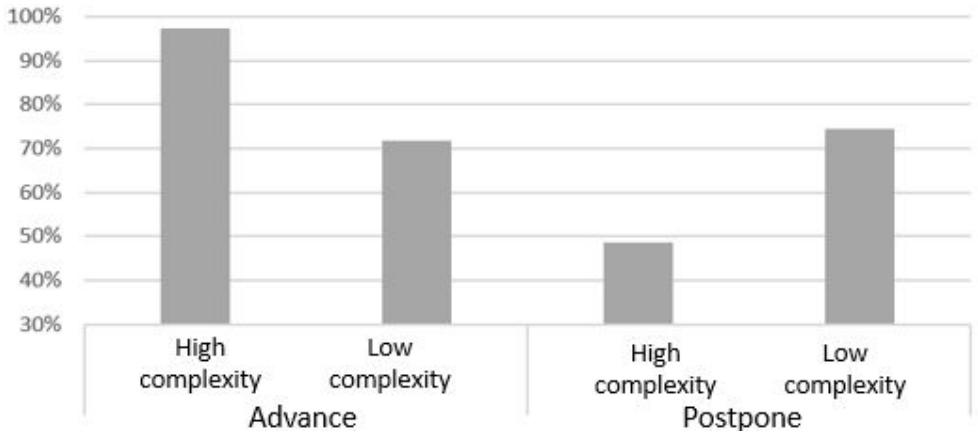

Figure 5: Compliance with AI proposition for total transparency, high/low complexity and advancing (-30) or postponing (30) the operation.

Those primary results tend to indicate that risk transparency (Fig. 1) leads to less compliance than with reliability transparency (Fig. 2) or total transparency (Fig. 3). When we visually compare reliability transparency and total transparency, we can see that total transparency leads also to a less compliant behavior.

CONCLUSION

Unlike Chancey et al. (2017) those first results show that transparency on risk has an impact on compliance response from participants. When the AI is transparent on the possible outcomes of the situation, it leads participants to be less compliant with the AI.

Thus, this effect is not the same for all situations. This result can be linked with Hancok et al. (2011) that define context or situational as a factor in trust. The possible reason is that a high complex scenario and a proposition to advance (and its opposite, i.e., low complex scenario and postponing proposition) might cancel the transparency effect as the participant could chose to comply because:

- it's more complex but it reduces the danger to advance
- it's less complex therefore there is less danger to postpone

However, it is interesting to see that total transparency mitigates the effect of risk transparency. When AI is transparent about its reliability and the risk, the compliance is in between the two. Participants might have used both of the information.

More in-depth statistics are needed, using ordinal regression logistic in order to see if those preliminary results are significant. The next objectives are to see if there is a correlation between subjective trust, risk perception and compliance.

To conclude we can see that there seems to be a difference between two concepts that Chen et al. (2018) included in the same level of transparency (i.e., level three of Situation Awareness Transparency). Therefore, future experiments need to consider at a very fine level the transparency concept they are using as some other researchers are suggesting (Andrada, 2022). There is a need to understand very specifically the interaction between the different possible aspects of the transparency concept and their implication in the Human-Machine interaction.

ACKNOWLEDGMENT

The research presented in this paper is carried out in the context of the SEANATIC Project (N°2082C0023). This project is supported by the Future Investments Program (PIA) operated by ADEME (the French Environment and Energy Management Agency).

REFERENCES

Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. AI & SOCIETY. https://doi.org/10.1007/s00146-021-01326-6

Chancey, E., Bliss, J., Yamani, Y., & Handley, H. (2017). Trust and the Compliance-Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence. Human Factors The Journal of the Human Factors and Ergonomics Society, 59, 333–345. https://doi.org/10.1177/0018720816682648

Chen, J., Mishler, S., & Hu, B. (2021). Automation Error Type and Methods of Communicating Automation Reliability Affect Trust and Performance: An Empirical Study in the Cyber Domain. IEEE Transactions on Human-Machine Systems.

Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science, 19(3), 259–282. https://doi.org/10.1080/1463922X.2017.1315750

Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (2021). Can You Trust Your Trust Measure? Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 92–100. https://doi.org/10.1145/3434073.3444677

de Visser, E., Peeters, M. M. M., Jung, M., Kohn, S., Shaw, T., Pak, R., & Neerincx, M. (2020). Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams. International Journal of Social Robotics, 12. https://doi.org/10.1007/s12369-019-00596-x

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. International Journal of Human–Computer Interaction, 35(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Hancock, P. A., Billings, D. R., Schaefer, K., Chen, J., Visser, E. D., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. Hum. Factors. https://doi.org/10.1177/0018720811417254

Hart, S., G., & Staveland, L., E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology, 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in Automation. IEEE Intelligent Systems, 28(1), 84–88. https://doi.org/10.1109/MIS.2013.24

Jessup, S. A. (2018). Measurement of the Propensity to Trust Automation. 65.

Lyons, J. B. (2013, mars 15). Being Transparent about Transparency: A Model for Human-Robot Interaction. 2013 AAAI Spring Symposium Series. 2013 AAAI Spring Symposium Series. https://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5712

Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human–machine trust: A multi-study look at the perfect automation schema. Theoretical Issues in Ergonomics Science, 20(4), 440–458.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. Academy of Management Review, 20(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors: The Journal of the Human Factors and Ergonomics Society, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Roundtree, K. A., Goodrich, M. A., & Adams, J. A. (2019). Transparency: Transitioning From Human–Machine Systems to Human-Swarm Systems. Journal of Cognitive Engineering and Decision Making, 13(3), 171–195. https://doi.org/10.1177/1555343419842776

Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., & Gorecky, D. (2016). Towards an operator 4.0 typology: A human-centric perspective on the fourth industrial revolution technologies. proceedings of the international conference on computers and industrial engineering (CIE46), Tianjin, China, 29–31.

Wang, N., Pynadath, D. V., & Hill, S. G. (2016). Trust calibration within a human-robot team: Comparing automatically generated explanations. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 109–116. https://doi.org/10.1109/HRI.2016.7451741

Wilson, R., Zwickle, A., & Walpole, H. (2018). Developing a Broadly Applicable Measure of Risk Perception. Risk Analysis, 39. https://doi.org/10.1111/risa.13207

Zhang, D. C., Highhouse, S., & Nye, C. D. (2019). Development and validation of the General Risk Propensity Scale (GRiPS). Journal of Behavioral Decision Making, 32(2), 152–167. https://doi.org/10.1002/bdm.2102