On Autonomy and Meaning in Human-Automation Interaction

Shadan Sadeghian[1,*,†], Marc Hassenzahl[1,†]

[1]University of Siegen, Germany

Abstract

Progress in automation will change the way we use and experience technology. The complex, unpredictable, and proactive nature of these technologies will transform them from tools controlled by humans to counterparts in dialogue with them. The existing models and frameworks on human-automation interaction (HAI), however, are insufficient to define this transition. They are mostly performance-oriented and neglect the experiential and meaning-oriented aspects of HAI, or highly rely on principles of direct manipulation limiting the affordances of automation. This creates a demand for alternative interaction paradigms based on quasi-social interaction. This paper elaborates on the problems of the existing frameworks and hints at new forms of interaction that consider experiential and performance goals.

Keywords

Interaction with Automation, Human-AI Interaction, Automated systems

1. The Rise of Automation

Automation plays a crucial role in our lives. It subsumes a broad range of technologies able to autonomously carry out tasks formerly done by humans [1]. While current automation largely addresses pre-programmed routine tasks, advances in machine learning, natural language processing and computer vision, will lead to more adaptive, "intelligent" automation, able to replace intellectual, creative, and non-routine work (i.e., knowledge work) [2]. This will change the way we use and experience technology. Automated systems differ from conventional computational artifacts, typically controlled by humans through forms of direct manipulation, such as Graphical User Interfaces (GUI), Tangible Computing, and in recent years embodied and ubiquitous computing. For example, automated vehicles, perceive the world through sensors, learn, and act in a proactive and autonomous manner. In fact, they are in dialogue with humans. People cooperate with them rather than use them. This shifts the perception of systems from a passive extension of self to active counterparts or collaborators [3].

Automation, however, does not happen overnight. It evolves continuously. The original idea of an automated system had been to relieve humans entirely from performing unwanted or dangerous tasks. This, however, is barely the case. More realistically, various forms of automation shape novel types of human-computer interaction in which humans and automated systems collaborate in different ways. This change raised concerns regarding the autonomy, trust, accountability, acceptance, and performance (effectiveness and efficiency in achieving task goals) of such "teams" of people and machines in the field of human-computer interaction (HCI). A widespread shift from the notion of controlling computational systems to collaborating with them will inevitably change the role of humans and consequently their experience of enjoyment and meaning in use. For example, feelings of autonomy as a source of well-being might suffer when delegating tasks rather than doing them. To this end, a big future challenge for HCI is to design meaningful interaction with automated systems while maintaining performance goals. While HCI has extensive knowledge about how to design for control, far less is known about appropriate interaction paradigms for automated systems perceived as counterparts.

2. Human-Automation Interaction Models

In the last decades, several models and frameworks are proposed that describe the interaction (or allocation of tasks) between human and automated systems. Most of these models focus on the level of system autonomy in fulfilling task goals. Among these, the model by Parasuraman et al. [1] is one of the most recognized ones that applies a who-centered approach for allocating tasks between human and automation in 10 levels. This model was later extended to 12 levels by Endsley [4]. These models of levels of automation (LOA), are originally developed in the context of decision-making in safety-critical systems, and their function allocation focuses on perfor-

AutomationXP23: Intervening, Teaming, Delegating - Creating Engaging Automation Experiences, April 23rd, Hamburg, Germany

*Corresponding author.

†These authors contributed equally.

shadan.sadeghian@uni-siegen.de (S. Sadeghian); marc.hassenzahl@uni-siegen.de (M. Hassenzahl)

0000-0002-8590-656X (S. Sadeghian); 0000-0001-9798-1762 (M. Hassenzahl)

mance metrics [5]. With increasing levels of automation (LOA), however, the relationship between human and automated systems changes [6]. At the higher levels, humans are rather in dialogue with algorithmic, self-learning, self-reliant, and proactive systems than acting through them. Personal assistants, chatbots, conversational interfaces, and autonomous or even anthropomorphic robots are examples of such computational artifacts. They tend to be perceived by their users as autonomous, semi-intelligent agents, either incidentally because the reasons for their actions are opaque (e.g., in the case of complex simulations or deep learning algorithms), or deliberately because they are designed that way (e.g., in the case of anthropomorphic robots). This fundamentally impacts the relationship between humans and computational artifacts and suggests an alterity relationship, where technology becomes "other", i.e., a counterpart technology [7, p. 99]. The change of relationship leads to interaction paradigms in which the automation can have roles with different goals from the human and could even supervise them (e.g., teaching robots). Thus, in these systems, using only LOAs for defining interaction between human and automation is insufficient [5].

In his recent book **Human-centered AI**, Ben Shneiderman, criticizes the LOA model: *"[...]increases in automation must come at the cost of lowering human control. But this zero-sum assumption limits thinking about ways to increase human control and the level of automation. There is a better way"* [8, p. 48]. To ensure human control while increasing automation, he suggests a two-dimensional framework for human-centered AI (HCAI) with two axes for human control and automation. The framework strongly emphasizes on supervisory control in interaction with automation (AI) in which the human is primarily assigned to the monitoring of automation, and occasionally intervenes to overwrite decisions, address errors, or unexpected circumstances [9, 10]. While this form of interaction is proposed to ensure reliability, previous research has shown that humans are bad at monitoring and vigilance due to reasons such as boredom, fatigue, distraction, lack of situational awareness, and deskilling [11, 12]. A very well-known example of such a situation is the take-over situation in highly automated cars. This situation requires the "driver" of the automated vehicle to take over control when the driving automation reaches its limits. To perform a safe maneuver, the "driver" needs to shift his/her attention to the driving context, gain situational awareness, have the required skills to operate the car, and perform the right maneuver in a few seconds[13, 14]. This example shows that human supervision over automation not only is insufficient to ensure safety or reliability but also leads to one of the *Ironies of Automation* [15].

3. Experiential Aspects

The LOA model was originally designed based on static task allocation. This is done by either, allocating tasks to the better performer between human and automation by comparison, automating everything possible and allocating the rest to the human, or allocating based on cost and performance benefits [16]. Later, the concept of *"Adaptive Automation"* was introduced that allows more flexibility in task allocation to increase performance [17]. The HCAI framework also follows adaptive allocation. However, the performance-oriented perspective focuses primarily on the functional capabilities of the technology when designing human-automation interaction (HAI). In its purest form, this leads to a so-called left-over allocation of tasks to humans based on what the machine is unable to do reliably (e.g., [18]). Even if the distribution of tasks is more considerate, it mostly follows the principle of optimizing performance by, for example, designing automation in accordance with the cognitive abilities of humans (e.g., [1]).

From the human perspective, this is a severely limited approach. In a meta-analysis, Onnasch et al. [19] showed that while the automation of routine tasks increases the performance of a human-automation system, it has only a small effect on the experienced workload of the humans involved. Increasing levels of automation reduce experiences of workload but lead to a decrease in situational awareness and failure performance. Humans start to lack insight into the tasks performed and become less and less engaged. In the long-term, increasing levels of automation, thus, lead to deskilling and technological "illiteracy", as well as a sense of not contributing and, consequently, to the diffusion or even abandonment of responsibility [20]. This reveals a fundamental dilemma of automated systems: On their way to full automation, respective systems do not create more meaningful, relaxing, and augmenting work experiences for the people involved, but disengagement and increased stress. In her seminal work on the "Ironies of Automation", Bainbridge already discussed how the notion of replacing the human through automation may actually lead to more and not fewer problems [15]. These problems remain important when designing human-automation interaction (HAI) (e.g., [21]), nevertheless, the actual design challenges are more comprehensive than this: HAI is not only a cognitive problem but an affective-motivational problem on an individual and societal level. Despite this, research on the experience of meaningful and fulfilling interaction with autonomous systems and how to design for it is rare [22].

One way to shape meaningful positive experiences in interaction with AI is through fulfilling human psychological needs. Previous research has widely discussed the role of universal human needs in shaping positive experiences. The Self-determination Theory specifies the three needs for autonomy, relatedness, and competence as essential substances for individual well-being [23]. Later Sheldon et al. [24] extended this to a list of 10 psychological needs. Drawn on these works, in the context of HCI, Hassenzahl et al. [25] showed that fulfillment of psychological needs leads to positive experiences in interaction with technology. While the HCAI model aims to put the human in the center, it only does so by addressing the need for autonomy, which comes at the cost of limiting automation. In this regard, a recent work on a collaboration with robots at work Smids et al. [26] mentioned five characteristics of meaningful work: pursuing a purpose, social relationships, exercising skills and self-development, self-esteem and recognition, and autonomy. These characteristics are in line with the needs for relatedness, competence, popularity, security, and autonomy, respectively [27]. Another approach is through interactive (contrary to supervisory) allocation, and assignment of tasks between humans and automation according to human psychological needs. Examples of this approach are ”Hotzenplotz” by Klapperich and Hassenzahl [28], and its later adaptation by Frison et al. [29] for automated vehicles, which propose interaction concepts with automation that aim to fulfill the needs for competence, autonomy, and stimulation while maintaining automation advantages.

4. From a Tool to a Counterpart

Human-automation interaction (HAI) fundamentally differs from traditional, cognitively oriented human-computer interaction (HCI). In HCI, the human is the center of the design. Widespread interaction paradigms, such as direct manipulation [30], gesture-based interaction [31] or tangible computing [32] revolve around human action, and agency and seek to extend the user’s control over the environment through computational tools. Automation is different. It has its own, however, restricted, agency, which is often opaque and complex. Humans do not directly act through automation, but indirectly by instructing, supervising, or supporting it. Thus, HAI has the character of a co-performance of two more or less autonomous entities, the machine and the human [33]. This has fundamental consequences: Autonomous systems are not simply integrated elements of human practices. In contrast, they have their own practices, which are entangled with the practices of humans, but not identical. To give an example: while a typewriter will shape the writing practice of a human author (for example, compared to writing by hand), the author will never have the feeling that the typewriter actually writes her book. In contrast, an autonomous typewriter would have its own writing practice. The human author can, for example, take the output of the autonomous typewriter and further edit it to make it into her own; however, editing is not writing. While both practices intersect, they are not integrated. In fact, through the design of autonomous systems, humans are excluded from any new practices performed by them. Consequently, humans might not feel responsible for the output of autonomous systems and experience drastic changes in the meaning of their work. By introducing an autonomous typewriter, the author’s work does not become easier but drastically changes its nature.

Through their agency, opaqueness, and complexity, autonomous systems can appear to humans as counterparts rather than as tool-like extensions of their self ([3]). Despite being machines, people now tend to use social metaphors in interaction, such as delegating, collaborating, being in a team, or trusting. Both aspects, that is, the entangled but separate practices as well as perceiving autonomous systems as counterparts, lead to a focus on the particular relationships established between the human and the system. These relationships, in turn, are established through the roles assigned to the system. Is it an expert, apprentice, team member, subordinate, superior, or something completely different? In this view, interaction paradigms for autonomous systems and resulting experiences are inevitably social in nature and are heavily shaped by the relationships established through design. This has also been observed in previous studies (e.g., [34]) that demonstrate the pleasure and meaning people derive from driving, the direct control over the car, and a feeling of being “one” with it. The very same study also shows that already simple automation, namely an adaptive cruise control, created feelings of the car as other, which was experienced as supportive but also diminished the meaning derived from driving. In this case, driving even remained an activity, and was only lightly supported by automation. Nevertheless, its meaning changed considerably. It seems out of question, that more automation towards, for example, self-driving cars will inevitably lead to perceptions of the car as other.

Interestingly, while interaction paradigms based on the notion of technology as an extension of the self are highly developed and validated (e.g., direct manipulation, [35]), the insights into the quasi-social interaction paradigms with automated counterparts remain vague. Current HCAI approaches tend to underplay the changes in people’s relationships with technologies. For example, Shneiderman [8, p. 55] mentions that computers are not teammates, collaborators, or co-active partners, as many suggest *“[...] Humans are responsible for actions of the technology assists that they use.”* He is actually demanding to design HAI as empowering extension of the self, thereby perpetuating principles of direct manipulation and disregarding the challenges interaction with more and more autonomous systems will bring. This has been also discussed widely in panels/duels he had with Pattie Maes [36, 37], where she mentions *"some tasks I may just not want to do myself even if the interface was perfect. If my car had a perfect interface for fixing the engine, I still wouldn't fix it. I just don't want to bother with fixing cars. I want someone else to do it"* highlighting the importance of considering new paradigms for collaborating with, or delegating tasks to automation.

5. Machine Autonomy ≠ Anthropomorphism

An often-used approach in exploring the interaction between human and automated systems and the distribution of their tasks is comparing these two interaction partners. Concepts such as "men are better at, machines are better at'' (MABA-MABA) have been around since the 1950s [38]. Furthermore, the common notion of automation as a means to replace human actions leads the existing interaction paradigms for automated systems to mainly copy natural ways of interacting with humans. This is why some computational artifacts, such as social robots, deliberately prompt social metaphors through anthropomorphism. Those *"anthropomorphic [interaction] paradigms [...] augment the functionality and behavioral characteristics of a robot [...] that we can relate to and rationalize its actions with greater ease"*[39]. This has clear advantages since it transfers already existing knowledge and skills from human-human interaction to human-technology interaction [40]. However, the delegation of tasks to another "human" might lead to feelings of losing own autonomy and can be detrimental when expectations about the communication and interaction capabilities of the automated system cannot be fulfilled. This implies that, just because technology feels like a counterpart, we should not blindly anthropomorphize its interaction with us. Contrarily, we need to recognize its non-human characters and shape and define it as an "Otherware"–a counterpart with quasi-social relationships with us [3]. This is not limited to accepting that automation may be best framed as an otherware, with its own goals and practices, but needs to be reflected in the interaction paradigms we provide. Therefore, the standard of direct manipulation simply makes no sense here, since others are not to be direct-manipulated (e.g., you do not direct-manipulate your colleague). Thus, we need to accept the pseudo-social nature of interacting with otherware, and the altered relationship to technology it implies.

6. Conclusion

Good Human-Automation Interaction is an imminent, unsolved problem, despite the longstanding research into automation, its design, and effects. One problem is the inherent conflict between providing control to the human and the very fact that automation largely controls itself. Models of human-centered automation, provide no real solution to this. They simply demand to limit the autonomy of automation, and its power, in a way that it remains controllable by a human. While this is an understandable demand, it does seem a very productive approach, and we doubt it will be successful merely by insisting on limiting automation. Another approach is to make automation more transparent. While applaudable, this approach seems doomed as well. True control requires a deep understanding of what is going on. While a system may be able to better explain certain situations in hindsight, during the process itself, explaining everything will most likely lead to information overload on behalf of the human. The "transparent" automation may rather foster a pseudo-accountability, which we find deeply disturbing and unethical. We need more refined interaction paradigms here that will certainly be hybrid in using essentially social forms of interaction but doing so in a reinterpreted, machine-like way. While this is far from settled, it seems important that we do not neglect the experiential view of human-computer interaction. Whether direct manipulating computational tools or quasi-socially working with autonomous systems, the provided arrangements should be meaningful to the humans involved. This is certainly a challenge. From the perspective of psychological needs, the experience of autonomy will change. While paramount and out of question in direct manipulation paradigms, interaction with otherware will entail negotiating autonomies rather. Direct manipulation is also bound to foster competence experiences, which may become more difficult if it starts to get more unclear whether a work result can be attributed to the human or the machine. In contrast, the quasi-social nature of otherware may lead to new forms of relatedness, the experience of mutual support, shared values, and goals. Nowadays, all this remains largely unexplored. And we fear that this might remain so if the HAI and HCAI community does not acknowledge the fundamental difference between interacting with computational tools as we know them and highly autonomous systems.

References

[1] R. Parasuraman, T. B. Sheridan, C. D. Wickens, A model for types and levels of human interaction with automation, IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans 30 (2000) 286–297.

[2] C. Coombs, D. Hislop, S. K. Taneva, S. Barnard, The strategic impacts of intelligent automation for knowledge and service work: An interdisciplinary review, The Journal of Strategic Information Systems 29 (2020) 101600.

[3] M. Hassenzahl, J. Borchers, S. Boll, A. R.-v. d. Pütten, V. Wulf, Otherware: How to best interact with autonomous systems, Interactions 28 (2020) 54–57.

[4] M. R. Endsley, From here to autonomy: lessons learned from human–automation research, Human factors 59 (2017) 5–27.

[5] H. A. Abbass, Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust, Cognitive Computation 11 (2019) 159–171.

[6] J. Scholtz, Theory and evaluation of human robot interactions, in: 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, IEEE, 2003, pp. 10–pp.

[7] D. Ihde, Technology and the lifeworld: From garden to earth (1990).

[8] B. Shneiderman, Human-Centered AI, Oxford University Press, 2022.

[9] T. B. Sheridan, Telerobotics, automation, and human supervisory control, MIT press, 1992.

[10] G. A. León, E. K. Chiou, A. Wilkins, Accountability increases resource sharing: Effects of accountability on human and ai system performance, International Journal of Human–Computer Interaction 37 (2021) 434–444.

[11] J. S. Warm, W. N. Dember, P. A. Hancock, Vigilance and workload in automated systems, in: Automation and human performance: Theory and applications, CRC Press, 2018, pp. 183–200.

[12] W. C. Harris, P. A. Hancock, E. J. Arthur, J. Caird, Performance, workload, and fatigue changes associated with automation, The International Journal of Aviation Psychology 5 (1995) 169–185.

[13] P. Wintersberger, C. Schartmüller, S. Shadeghian-Borojeni, A.-K. Frison, A. Riener, Evaluation of imminent take-over requests with real automation on a test track, Human Factors 0 (0) 00187208211051435. URL: https://doi.org/10.1177/00187208211051435. doi:10.1177/00187208211051435. arXiv:https://doi.org/10.1177/00187208211051435, pMID: 34911393.

[14] S. Sadeghian Borojeni, L. Weber, W. Heuten, S. Boll, From reading to driving: priming mobile users for take-over situations in highly automated driving, in: Proceedings of the 20th international conference on human-computer interaction with mobile devices and services, 2018, pp. 1–12.

[15] L. Bainbridge, Ironies of automation, in: Analysis, design and evaluation of man–machine systems, Elsevier, 1983, pp. 129–135.

[16] W. B. Rouse, W. B. Rouse, Design for success: A human-centered approach to designing successful products and systems, volume 2, Wiley-Interscience, 1991.

[17] R. Parasuraman, T. Bahri, J. E. Deaton, J. G. Morrison, M. Barnes, Theory and design of adaptive automation in aviation systems, Technical Report, Catholic Univ of America Washington DC cognitive science lab, 1992.

[18] E. Hollnagel, Cognitive ergonomics: it's all in the mind, Ergonomics 40 (1997) 1170–1182.

[19] L. Onnasch, C. D. Wickens, H. Li, D. Manzey, Human performance consequences of stages and levels of automation: An integrated meta-analysis, Human factors 56 (2014) 476–488.

[20] R. Parasuraman, T. B. Sheridan, C. D. Wickens, Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs, Journal of cognitive engineering and decision making 2 (2008) 140–160.

[21] T. Hancke, Ironies of automation 4.0, IFAC-PapersOnLine 53 (2020) 17463–17468.

[22] V. Roto, P. Palanque, H. Karvonen, Engaging automation at work–a literature review, in: IFIP Working Conference on Human Work Interaction Design, Springer, 2018, pp. 158–172.

[23] R. M. Ryan, E. L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being., American psychologist 55 (2000) 68.

[24] K. M. Sheldon, A. J. Elliot, Y. Kim, T. Kasser, What is satisfying about satisfying events? testing 10 candidate psychological needs., Journal of personality and social psychology 80 (2001) 325.

[25] M. Hassenzahl, S. Diefenbach, A. Göritz, Needs, affect, and interactive products – Facets of user experience, Interacting with Computers 22 (2010) 353–362. URL: https://doi.org/10.1016/j.intcom.2010.04.002. doi:10.1016/j.intcom.2010.04.002. arXiv:https://academic.oup.com/iwc/article-pdf/22/5/353/1997205/iwc22-0353.pdf.

[26] J. Smids, S. Nyholm, H. Berkers, Robots in the workplace: a threat to—or opportunity for—meaningful work?, Philosophy & Technology 33 (2020) 503–522.

[27] S. Sadeghian, M. Hassenzahl, The "artificial" colleague: Evaluation of work satisfaction in collabo-

ration with non-human coworkers, in: 27th International Conference on Intelligent User Interfaces, IUI '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 27–35. URL: https://doi.org/10.1145/3490099.3511128. doi:10.1145/3490099.3511128.

[28] H. Klapperich, M. Hassenzahl, Hotzenplotz: reconciling automation with experience, in: Proceedings of the 9th Nordic conference on human-computer interaction, 2016, pp. 1–10.

[29] A.-K. Frison, P. Wintersberger, A. Riener, C. Schartmüller, Driving hotzenplotz: A hybrid interface for vehicle control aiming to maximize pleasure in highway driving, in: Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications, 2017, pp. 236–244.

[30] E. L. Hutchins, J. D. Hollan, D. A. Norman, Direct manipulation interfaces, Human–computer interaction 1 (1985) 311–338.

[31] L. Gupta, S. Ma, Gesture-based interaction and communication: automated classification of hand gesture contours, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 31 (2001) 114–120.

[32] E. Hornecker, J. Buur, Getting a grip on tangible interaction: a framework on physical space and social interaction, in: Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006, pp. 437–446.

[33] L. Kuijer, Automated artefacts as co-performers of social practices: washing machines, laundering and design, in: Social Practices and Dynamic Non-Humans, Springer, 2019, pp. 193–214.

[34] K. Eckoldt, M. Knobel, M. Hassenzahl, J. Schumann, An experiential perspective on advanced driver assistance systems (2012).

[35] B. Shneiderman, The future of interactive systems and the emergence of direct manipulation, Behaviour & Information Technology 1 (1982) 237–256.

[36] B. Shneiderman, P. Maes, Direct manipulation vs. interface agents, Interactions 4 (1997) 42–61. URL: https://doi.org/10.1145/267505.267514. doi:10.1145/267505.267514.

[37] D. Wang, P. Maes, X. Ren, B. Shneiderman, Y. Shi, Q. Wang, Designing ai to work with or for people?, in: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–5.

[38] P. M. Fitts, Human engineering for an effective air-navigation and traffic-control system. (1951).

[39] B. R. Duffy, Anthropomorphism and the social robot, Robotics and autonomous systems 42 (2003) 177–190.

[40] L. Damiano, P. Dumouchel, Anthropomorphism in human–robot co-evolution, Frontiers in psychology 9 (2018) 468.