OPEN FORUM

Blue collar with tie: a human-centered reformulation of the ironies of automation

Norman Meisinger[1]

Abstract

When Lisanne Bainbridge wrote about counterintuitive consequences of the increasing human–machine interaction, she concentrated on the resulting issues for system performance, stability, and safety. Now, decades later, however, the automized work environment is substantially more pervasive, sophisticated, and interactive. Current advances in machine learning technologies reshape the value, meaning, and future of the human workforce. While the 'human factor' still challenges automation system architects, inconspicuously new ironic settings have evolved that only become distinctly evident from a human-centered perspective. This brief essay discusses the role of the human workforce in human–machine interaction as machine learning continues to improve, and it points to the counterintuitive insight that although the demand for blue-collar workers may decrease, exactly this labor class increasingly enters more privileged working domains and establishes itself henceforth as 'blue collar with tie.'

Keywords Automation · Ironies · Human–machine interaction · Human factor · AI · Interactive machine learning · Machine teaching · Human–machine systems

1 Introduction

Almost 40 years ago, Bainbridge (1983) reflected on *ironies of automation*. Her brief but astute and strongly referenced paper oscillates around the fact that machines work more precisely and more reliably than their operators, although "the more advanced a control system is, so the more crucial may be the contribution of the human operator" (ibid., p. 775) in case of anomalies.

Whereas Bainbridge, and many others recently, including Strauch (2018), elaborated on the ironic consequences for system performance, stability, and safety inherent to socio-technical systems, a human-centered reformulation would address inevitable consequences for human operators that face a constantly rising automized world and ironically rise in number as well, although machines are increasingly replacing them. Their number may not grow in absolute terms, but in relative terms by gradually shaping a new working class that can be labeled 'blue collar with tie.'

While machines initially served only a passive, i.e., physical and cognitive supporting, function in technical work environments, nowadays machines predict (Gill 2020), recommend (Milano et al. 2020), create artistically (Elgammal et al. 2017), and even decide autonomously (Héder 2020)[1] throughout crucial instances of value creation processes. Machines are more and more seamlessly coupled to each other and even learn automatically without human intervention. Such artificial workforces progressively outperform the human workforce in a variety of ways. From a broader perspective, the discourse should, therefore, not only belong to human–machine *interaction* but also human–machine *substitution*. In other words, an essential tipping point occurs: the assistance of humans by machines turns into the assistance of machines by humans. The structural shift that has occurred between Bainbridge's analysis and the present inquiry—with Collins (2021, p. 59), humankind passed level I ("engineered intelligence") and arrived already at level II, where the machine "does the job that a human once did"—resulted in new ironic 'human-centered' consequences.

✉ Norman Meisinger
n.meisinger@zeppelin-university.net

1 Department of Business & Economics, Zeppelin Universität Friedrichshafen, Fallenbrunnen 3, 88045 Friedrichshafen, Germany

---

[1] When evolutionary algorithms are involved, the machines' operations may not even be entirely reconstructable (Wilson et al. 2018).

However, 'human-centered' does not mean that her approach neglected human factors—on the contrary, precisely that was her hook—but "places human needs, purpose, skill, creativity, and human potential at the center of activities of human organisations and the design of technological systems" (Gill 1996, p. 110).

After a brief and general outline of Bainbridge's core idea, each of the three ironic facets will be introduced as originally discussed, followed by a concise human-centered reformulation that considers intermediate developments.

2 The ironies revisited

Bainbridge (1983, p. 775) conceived *irony* as a "combination of circumstances, the result of which is the direct opposite of what might be expected." The starting point of her investigation was the operator. The operator is laden "with responsibility for abnormal conditions" (ibid.) but should nevertheless "be eliminated from the system" because from the designer's perspective, "the operator is unreliable and inefficient," notwithstanding the fact that the designer's own "errors can be a major source of operating problems" (ibid.). Ironically "the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate" (ibid.).

Given that an operator's intervention can even have a negative impact on system performance, Bainbridge elaborated on ironic problems in terms of (1) *knowledge* and *skills*, (2) *monitoring*, and (3) the operator's *attitude*. For all three, *meanwhile*, corresponding human-centered ironies can be uncovered as well.

2.1 Skills and knowledge

First, Bainbridge (1983, p. 775) recognized that in the course of the decreasing variety and frequency of human activities, the operators' "physical skills deteriorate when they are not used, particularly the refinements of gain and timing. This means that a formerly experienced operator who has been monitoring an automated process may now be an inexperienced one" although these skills would exactly be needed in case of system anomalies. Further she recognized that "an operator will only be able to generate successful new strategies for unusual situations if he has an adequate knowledge of the process" (ibid.). Both ironical reflections exemplify the contradictory nature of tasks '*after*' automation and refer to the declining human performance as a product of unlearning.

Meanwhile, however, the operator is hardly ever required for the firefighter's role beyond automations' abilities. Procedures for system anomalies have been highly professionalized with technical resources, such as dissimilar redundancies, intelligent watchdog routines that go ahead of simple dead man's switches, and preventive organizational methods, such as 'failure mode and effects analysis,' or 'six sigma.' Operators increasingly perform repetitive tasks that either require a certain kind of knowledge that cannot (yet) be codified or tasks that regularly alternate as per unpredictable inputs (new orders, raw material quality) to such an extent that a replacement by machines would be unprofitable (for instance, machine setups, or variant diversity).

However, as with the machines' more progressive pervasion of social spheres and its increasing attraction for business, recently AI experts recognized that machine learning's economization potential and ever new applications develop over-proportionally to the number of experts that are able to continue advances in machine learning correspondingly. Tens of thousands of machine learning scientists and even hundreds of thousands of data analysts are far out of proportion to tens of millions of domain experts (Simard et al. 2017). The gradually emerging technology of *interactive machine learning* (or in turn: *machine teaching*, i.e., when machines are taught by their operators) promises a new business case with tremendous potential: "By enabling domain experts to teach, we will enable them to apply their knowledge to solve directly millions of meaningful, personal, shared, 'one-off' and recurrent problems at a scale that we have never seen" (ibid., p. 10–11).

So far, machine learning primarily extracts knowledge from (preferably) large *data sets* in mutual complementation with human and organizational learning. Interactive machine learning, however, extracts knowledge and thus learns no more only from data sets but primarily from *workers* and *users*. The millions of human workforces which interact with machines bear a vast amount of unexploited and valuable domain knowledge that has been gained and refined over their entire working careers but was at least so far non-extractable.

The irony now is that operators are expected to help the machine reduce their own value score in a break-even calculation against its substitution by machines. Compared to machines, operators had always borne in the burden of proof for its necessity. Costs associated with the purchase of a machine, on the other hand, are meticulously calculated. After acquisition, the machine's specificity, the frontload spent, its high initial loss of market value, and the adapted framework conditions, ultimately the resulting path dependency (Sydow et al. 2009), are far more pronounced than that of the (more or less) low-loss, widely available operators. Even more obscure: while at Bainbridge's time the operator in the role of a firefighter allowed the machine as such to be used at all, today, ironically, the machine increasingly supplants him—even by his assistance.

Nowadays, operators share the given labor no longer with human coworkers but rather with artificial entities (robotics,

algorithms) as their 'new colleagues.' However, these new entities come with different teamwork and knowledge-sharing principles, as the typical human colleagues may. As the operator's new 'colleagues,' the taught machines, however, do not share the new knowledge they have acquired. The learning direction between operators and machines may even be entirely unipolar. The machine learns from the operators, but not the operator from the machines. Thus, *ironically* the machine gradually rationalizes its colleague even with his own help.

2.2 Monitoring

Second, Bainbridge (1983, p. 776) recognized "that the automatic control system has been put in because it can do the job better than the operator, but yet the operator is being asked to monitor that it is working effectively." That means "if the decisions can be fully specified, then a computer can make them more quickly, taking into account more dimensions and using more accurately specified criteria than a human operator can. There is, therefore, no way in which the human operator can check in real-time that the computer is following its rules correctly" (ibid.). Additionally, "it is impossible for even a highly motivated human being to maintain effective visual attention towards a source of information on which very little happens, for more than about half an hour" (ibid.). These ironical reflections refer to the *monitoring* of automated processes, given a human's attention span, which declines as the automation's performance enhances.

Meanwhile, however, the monitoring direction is bidirectional (in some cases just reversed) even though knowledge, on the contrary, flows only from the operator to the machine. While several decades ago, machines processed work stringently rule by rule, nowadays they learn as well from their human coworkers. As the operator's 'new colleagues,' artificial entities extract knowledge that designers cannot express in rules by being taught and observing their performing human counterparts. Whether they want to or not, operators reveal a certain kind of their knowledge to the machines, ironically even if they cannot articulate it. Even though knowledge hiding is a widespread strategy to safeguard one's raison d'être in firms against rival human colleagues (Connelly et al. 2012), any strategy to hide knowledge would run into trouble with machines because the erroneous teaching of machines automatically leads to errors in one's own account.

Even though the sensual experience posed a USP of the qualified worker for a long time, as Brödner (1989) empirically elucidated in the case of the computer-aided craftsman 30 years ago, machines nowadays not only imitate the human senses but vastly surpass their performance. In 1966

Polanyi proposed to distinguish *knowing* according to its articulability.[2] In shaping the insight that "*we can know more than we can tell*" (Polanyi 1966, p. 4), he inspired numerous sociologists and philosophers of AI to explore how machines could access such tacit knowledge. Tacit knowledge materializes in practice (Brödner 2019, p. 206). Thus, it is—if not unspecifiable or even ineffable—at least "more complex to untangle than others" (Lowney 2011, p. 20), and can be seen as a part of human experience and action ever coupled to meaning that is "*not directly capable of being digitized*" (Gulick 2020).

In the latest machine teaching applications, the domain experts are in vivo "providing indicative samples, describing indicative features, or otherwise selecting high-level model parameters" (Dudley and Kristensson 2018, p. 4) that are charged with all their experiences and skills. With such machine learning technologies, where "human intelligence is applied through iterative teaching and model refinement in a relatively tight loop of 'set-and-check'" (ibid.), the value of human teachers gradually becomes obsolete by explicating their meticulously acquired experiential/empirical knowledge. Machines increasingly absorb diverse kinds of knowledge (albeit not yet craft skills) that humans cannot entirely explicate in diverse domains, such as failure detection at factory quality gates, NC calibration, digital pathology with microscopic images of tissue samples (Lindvall et al. 2018) and interpretation of radiographs, detecting errors in insurance claims (Ghani and Kumar 2011), or research citation screening for systematic medical reviews (Wallace et al. 2012).

Judged from a human-centered perspective, the value of human workforces in working environments gradually depletes given the fact that explicable and indeed explicated knowledge also accounts for differences in value and state of human workforces. Therefore, the crucial juncture of the present inquiry may not be the type of knowledge (ineffable, unspecifiable, tacit, explicit, or so),[3] but the fact that operators automatically (must) help crowd themselves out by simply doing for what they are paid and have been initially educated.

---

[2] Gulick (2020) vividly pointed out that this idea, albeit slightly modified, has many origins in other conceptions and disciplines as well (e.g., Ryle's "knowing how," Heidegger's "present at hand," Kahneman's "thinking fast").

[3] Even though this long-standing debate continues to gain attraction (Brödner 2019; Collins 2018; Gulick 2020; Lowney 2011), especially today, against the backdrop of the emerging technological field of interactive machine learning and machine teaching.

2.3 Operator attitudes

Third, Bainbridge (1983, p. 776) recognized that the "level of skill that a worker has is also a major aspect of his status, both within and outside the working community. If the job is 'deskilled' by being reduced to monitoring, this is difficult for the individuals involved to come to terms with. It also leads to the ironies of incongruous pay differentials, when the deskilled workers insist on a high pay level as the remaining symbol of a status which is no longer justified by the job content." Her ironical reflections refer to the operators' status preservation that becomes more difficult the more automized the firms' workflows are.

Meanwhile, however, no longer is only the typical operator threatened by machine rationalization. Nowadays, the human–machine substitution spreads as well to professions with higher education, such as the physician, and ironically even to the machines' designers themselves. Machines gradually slide as an exhausting mediation layer between the workers/experts and the actual activity they were once educated for. The reliability, learning, and performance of this layer have gained tremendous advances during the last decades.

One of the most frequently noted consequences of that progress is that low-skill and low-wage jobs are endangered to be outperformed and even replaced by machines (Frey and Osborne 2017). On the other hand, the applied perspective emphasizes an egalitarian devaluation of competencies from operator to designer. Interactive learning machines constitute an exhausting mediation layer that gradually converges the broad spectrum of labor classes that have previously been distinguished by skills and knowledge, scarcity, and value in the labor markets. Consequently, not only the worker becomes an operator. The medical doctor, however, becomes an operator as well, while the machine ironically becomes anthropomorphized—for instance, as a "person" with outperforming medicine expertise (Bunz and Braghieri 2021). Moreover, even the machine designer becomes an operator since the increasing open-source movement shares highly complex code modules that can be adapted and employed even though they may not be thoroughly understood.

3 Conclusion

The essay intends to raise awareness for the structural change caused by the increasing deployment of machines in work environments that occurred since Bainbridge's analysis. In work environments where humans and machines interact, 'blue collar with ties' obviously are becoming fashionable. The idea "that computerisation, automation and use of robotics devices will automatically free human beings from soul destroying backbreaking tasks and leave them free to engage in more creative work" still remains a myth (Gill 1996, p. 111). Human workers are subtly loaded with dilemmatic demands, competence erosions, status losses, and damage to work identity.

Data availability Not applicable.

Code availability Not applicable.

Material availability Not applicable.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states no conflict of interest.

References

Bainbridge L (1983) Ironies of automation. Automatica 19:775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Brödner P (1989) In search of the computer-aided craftsman. AI Soc 3:39–46. https://doi.org/10.1007/BF01892674

Brödner P (2019) Coping with Descartes' error in information systems. AI Soc 34:203–213. https://doi.org/10.1007/s00146-018-0798-8

Bunz M, Braghieri M (2021) The AI doctor will see you now: assessing the framing of AI in news coverage. AI Soc. https://doi.org/10.1007/s00146-021-01145-9

Collins H (2018) Artifictional intelligence: against humanity's surrender to computers. Polity Press, Medford

Collins H (2021) The science of artificial intelligence and its critics. Interdiscip Sci Rev 46:53–70. https://doi.org/10.1080/03080188.2020.1840821

Connelly CE, Zweig D, Webster J, Trougakos JP (2012) Knowledge hiding in organizations. J Organ Behav 33:64–88. https://doi.org/10.1002/job.737

Dudley JJ, Kristensson PO (2018) A review of user interface design for interactive machine learning. ACM Trans Interact Intell Syst 8:1–37. https://doi.org/10.1145/3185517

Elgammal A, Liu B, Elhoseiny M, Mazzone M (2017) CAN: creative adversarial networks, generating "art" by learning about styles and deviating from style norms. Arxiv (21 Jun 2017). http://arxiv.org/pdf/1706.07068v1

Frey CB, Osborne MA (2017) The future of employment: how susceptible are jobs to computerisation? Technol Forecast Soc Change 114:254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Ghani R, Kumar M (2011) Interactive learning for efficiently detecting errors in insurance claims. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, CA, USA, pp 325–333. https://doi.org/10.1145/2020408.2020463

Gill KS (1996) The human-centred movement: the British context. AI Soc 10:109–126. https://doi.org/10.1007/BF01205277

Gill KS (2020) Prediction paradigm: the human price of instrumentalism. AI Soc 35:509–517. https://doi.org/10.1007/s00146-020-01035-6

Gulick WB (2020) Machine and person: reconstructing Harry Collins's categories. AI Soc. https://doi.org/10.1007/s00146-020-01046-3

Héder M (2020) The epistemic opacity of autonomous systems and the ethical consequences. AI Soc. https://doi.org/10.1007/s00146-020-01024-9

Lindvall M, Molin J, Löwgren J (2018) From machine learning to machine teaching: the importance of UX. Interactions 25:52–57. https://doi.org/10.1145/3282860

Lowney C (2011) Ineffable, tacit, explicable, and explicit: qualifying knowledge in the age of "intelligent" machines. Tradit Discov Polanyi Soc Period 38:18–37. https://doi.org/10.5840/TRADDISC2011/20123819

Milano S, Taddeo M, Floridi L (2020) Recommender systems and their ethical challenges. AI Soc 35:957–967. https://doi.org/10.1007/s00146-020-00950-y

Polanyi M (1966) The tacit dimension. Routledge & Kegan Paul, London

Simard PY, Amershi S, Chickering DM, Pelton AE, Ghorashi S, Meek C, Ramos G, Suh J, Verwey J, Wang M, Wernsing J (2017) Machine teaching: a new paradigm for building machine learning systems. Arxiv (11 Aug 2017). http://arxiv.org/pdf/1707.06742v3

Strauch B (2018) Ironies of automation: still unresolved after all these years. IEEE Trans Hum Mach Syst 48:419–433. https://doi.org/10.1109/THMS.2017.2732506

Sydow J, Schreyögg G, Koch J (2009) Organizational path dependence: opening the black box. Acad Manag Rev 34:689–709. https://doi.org/10.5465/amr.34.4.zok689

Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA (2012) Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: Proceedings of the 2nd ACM sighit international health informatics symposium, Miami, FL, USA, pp 819–824. https://doi.org/10.1145/2110363.2110464

Wilson DG, Cussat-Blanc S, Luga H, Miller JF (2018) Evolving simple programs for playing Atari games. In: Proceedings of the genetic and evolutionary computation conference, July 15th–19th 2018, Kyoto, Japan, pp 229–236. https://doi.org/10.1145/3205455.3205578