THE MEANING OF 'GOOD DESIGN' IN THE AGE OF SMART AUTOMATION – OR, WHY HUMAN-CENTERED DESIGN NEEDS ETHICS

RODRIGO HERNÁNDEZ-RAMÍREZ

Universidade Europeia
IADE - Faculdade de Design, Tecnologia e Comunicação
UNIDCOM/IADE - Unidade de Investigação em Design e Comunicação
rodrigo.ramirez@universidadeeuropeia.pt

https://doi.org/10.34632/jsta.2020.9490

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114

ABSTRACT

The increasing adoption of smart automation has improved people's lives in several ways, but it has also brought a host of new problems such as deskilling, deepening of structural inequalities, new forms of exploitation, loss of privacy and hindering of human liberties. This paper begins by assuming that such issues are the consequence of poor design and takes the opportunity to analyse what "good design" should mean in turn. Following insights from philosophy of technology, it surveys the general inherent complexities of automation and argues that Human-Centered Design (HCD) endorses an instrumentalist conception of technology. This paper shows that such a conception of human–technology relations significantly limits designers' capacity to approach design from a genuinely ethical standpoint. The paper concludes with a sketch of principles that HCD should incorporate to become a truly humanist and ethically minded design approach. Ultimately, this paper argues that designers should seek not only to satisfy users' needs but actively reflect on the role that technologies play in helping humans to attain a "good life". In other words, they should start by asking themselves what our human project for the future is.

Keywords: Ethics; Instrumentalism; Mediation theory; Philosophy of design; Smart automation; User-Centered design.

1. INTRODUCTION

Over the past decades, thanks to a combination of new Artificial Intelligence (AI) methods and the exponential growth of computing power, Internet networks, and data availability, automation has become "smart". In most developed regions of the world, smart automation has turned vital for human well-being. A growing number of consumer and industrial products and services across every area of human activity now depend on some form of smart, automated system. At the same time, the growth of automation is bringing new forms of exploitation, discrimination, and abuse; deepening existing structural inequalities and severely threaten privacy and democratic liberties. Seen from a purely instrumentalist standpoint, many of these systems are well built and remarkably efficient. However, from a moral standpoint, they are nothing short of nefarious. Such poorly designed systems (in the sense that they worsen, rather than improve people's lives) can be attributed to incompetence, lack of resources, ignorance, misguided principles, or outright malice. Nevertheless, whatever the explanation, it is clear that the designers who developed these systems failed to consider the ethical consequences of their creations or, at best, treated them as an afterthought. This lack of explicit ethical guidelines, however, is not exclusive of misshapen design approaches, but a problem that also afflicts Human-Centered Design (HCD), the dominant paradigm in digital product design nowadays.

Avoiding poor design is always necessary, but in the case of smart automation the situation is urgent, for nowadays we are setting the pace and tone of the technological developments that will shape society for decades to come. That is why one of the central questions that people involved in the design of smart automation must address from the outset is how to create systems that genuinely improve people's lives while minimising potential negative impacts. In other words, they ought to think about what "good design" should mean in the age of smart automation.

This paper takes the above question as a starting point. Following insights from philosophy of design and philosophy of technology, this paper argues HCD should revise its implicit commitments to technological instrumentalism. This paper argues that to become genuinely humanistic, HCD should focus not merely on satisfying users' needs. It contends that designers should strive to make explicit their assumptions and biases about technological systems and the long-term impact they can have on people and society by focusing not on the technical and methodological obstacles impeding their fruition but on the minimum ethical requirements they should meet. Ultimately, this paper argues that design has a crucial responsibility in developing our human project for the future. It is important to note, however, that the principles here outlined are nothing but rough sketches and therefore still require considerable reworking before reaching a potential normative level. The exercise here developed should be seen as a starting point for developing a more thorough critique of contemporary design approaches and how they respond to the social challenges imposed by smart automation.

https://doi.org/10.34632/jsta.2020.9490

2. THE PROBLEMS WITH SMART AUTOMATION

Humans have been automating tasks for a long time; this propensity to "off-load" work to artificial systems[1] is arguably a crucial aspect of being human (Martinho-Truswell, 2018). Traditionally, automation involved delegating the execution of physical or non-cognitive tasks initially carried out by humans or animals (e.g., beasts of burden) to mechanical devices (Danaher, 2018; Johnson & Verdicchio, 2017). In the last decades, however, the development of computational tools in general, and AI in particular, has enabled automated systems to take over cognitive tasks. Since their aim is to replace "human manual control, planning and problem solving" (Bainbridge, 1983, p. 775) to carry out specific tasks, smart automated systems may be regarded as instances of "cognitive outsourcing" (Danaher, 2018).

Automation implies that a system is capable of accomplishing a given process without human intervention during runtime. Nonetheless, all types of automation still require human supervision; if not to start them, at least to guarantee their proper functioning, to adjust them, to provide maintenance, and to improve and expand their features (Bainbridge, 1983). For simple automated systems, the former tasks are less problematic; but as their complexity grows, so does the oversight they require. Since it is a human that ensures an automated system is operating safely and is also responsible for mitigating its potential malfunctions, it follows that the more complex the automation, the more crucial the role of the human becomes (Strauch, 2018). In this sense, and regardless of how sophisticated they might be, automated systems can still be regarded as human–machine systems (Bainbridge, 1983).

The main reasons for adopting automated systems are that they are faster, more reliable, and efficient than human beings carrying out the same tasks; this tends to be true for the most part. Ironically, however, errors or operational constraints introduced when automated systems are designed can become a source of problems that, invariably, need to be addressed by humans (Bainbridge, 1983; Strauch, 2018). Designing an automated system implies abstracting processes; breaking them down into definite tasks and then organising them into sequences. Hence, automation implies a deep understanding of the processes it takes over. Tasks that follow explicit procedural logic (e.g., simple arithmetic operations) are easier to formalise and automate. That is why machines vastly surpass human speed and accuracy when it comes to computing (Autor, 2014).

However, that is not the case for processes that require flexibility, judgement, common sense, and improvisation. Tasks such as cooking an omelette or inventing a (good) joke are easy for humans to accomplish, but difficult to automate; not because humans know the "rules" governing them, but because we quickly develop *tacit* understanding of how they work (Autor, 2014). Ironically, then, although systems are explicitly designed to replace human work, *other* humans end up carrying out the tasks that could not be automated (Bainbridge, 1983). That is why we still

1 A system may be understood as an entity that can be separated into parts, which are all simultaneously linked to each other in specific ways (Vermaas et al., 2011, ch. 5).

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114 https://doi.org/10.34632/jsta.2020.9490

have to make all sorts of input graspable for machines. To paraphrase the late anthropologist David Graeber (2018), the problem of automating tasks that are simple for humans but complicated for machines to accomplish is that users have to carry out essentially all the semantic labour. To put it in Flusserian terms, in this relation the human "functions as the functionary of a function" (Flusser, 2014).

Historically, automation has been more prevalent in sociotechnical systems[2] such as agriculture, industry, and transportation. Ubiquitous computing and new AI methods, however, have allowed smart automation to become integral for (and in the process radically change) a growing number of everyday human activities. Smart AI-powered automated systems require vast amounts of data, both to be developed (trained) and to carry out their tasks. Whereas "traditional" automation relied on a fixed set of instructions based on the designer's understanding of the process that would be automated, smart automation is more ambitious. For example, smart automation based on Machine Learning (ML) relies on patterns derived from vast amounts of data about human conduct. Usually, the services provided by smart automation require predicting human behaviour, so the more data they have about a user's context and activities, the better they perform[3]. Hence why smart, automated systems are often deliberately conceived to acquire as much data as possible from their users, usually without their explicit knowledge and consent.

User data has become so valuable that certain forms of automation are built for no other reason than to create a data surplus. This continuous extraction and accumulation of information is the backbone of a new exploitative economic system threatening fundamental human liberties called "surveillance capitalism" (Zuboff, 2015). Entire organisations are now invested in developing such "dishonest" forms of automation that not only extract data but, to do so, also force onto users behaviours for which they have no legitimate need or desire (Girardin, 2019). These and other (supposedly benign) forms of smart automation, such as algorithms supporting decision-making, can easily lead to new forms of exploitation, discrimination, and abuse; they can deepen existing structural inequalities and severely threaten privacy and democratic liberties.

Such poorly designed systems (in the sense that they worsen, rather than improve people's lives) can be attributed to incompetence, lack of resources, ignorance, misguided principles, or even malice. However, in the case of contemporary smart automation, the main culprit—besides the ruthless logic of growth imposed by venture capital—is arguably a mixture of overconfidence, cultural biases, and misunderstandings concerning human–technology relations. Technically minded people responsible for inadequate smart, automated systems tend to focus solely on the immediate positive aspects of technologies and scantly on their potential downsides; usually caring more for the economic success of their creations than for the well-being of their users (Norman, 2010). Rarely do they stop to ponder whether their creations should materialise in the first place. In other words, they fail to consider the ethical consequences of their designs. Generally, these designers fail to consider that any new

2 A sociotechnical system may be regarded as a hybrid system (i.e., one whose components belong to many different *worlds*) with an extremely high degree of complexity that has many users at any given moment, and that involves people both as users and operators (Vermaas et al., 2011).

3 The problem, however, is that ML models are often inscrutable, meaning that those who conceived and trained them do not have a clear understanding of exactly how they operate.

https://doi.org/10.34632/jsta.2020.9490

product or service will be embedded within larger and massively complex sociotechnical systems, leading to unforeseen consequences and risks for society as a whole. Such an approach to design that is not guided by (ethically grounded) human interests is not only poor but *wrong*. The question then is, what would be an acceptable way to develop less problematic smart automated systems. That is, what does *good* design should mean in the age of smart automation?

3. THE IDEA OF "GOOD DESIGN"

Asking what a good design approach is necessarily implies asking what good design is. Both are questions of an ethical nature. Ethics is understood here as the broader problem space of how people should lead their lives best and, therefore, what should we strive to be, and which activities are genuinely valuable to pursue, and which are not. Since most, if not all, of these activities often require interacting with some artificial (i.e., designed) thing, it is clear that design has a strong relation to ethical issues.

The idea of "good design" and that design has a powerful impact over human life is a tenet of this discipline since its origins in the late nineteenth century. The emergence of the Arts and Crafts Movement—arguably the cradle of modern theorisation about design—was driven by concerns about how the (good) quality of designed objects could affect not only individuals but societies at large. Modernism sought to strengthen this link, shaping the idea of design as a vehicle for human progress. By blurring the divide between ethics and aesthetics, the modernist understanding of good design assumed that artefacts that resulted from the proper synthesis of strictly necessary formal beauty and functionality would naturally improve people's values and behaviour (Buchanan, 2005). Modernist views on design cannot be detached from its broader views about art, politics, society, and human nature, to the contrary, they are intimately related (Parsons, 2015). The modernist conception of design, despite its excessive focus on functionality, continues to be a necessary reference, not only for understanding the evolution of the discipline, but also the intrinsically ethical nature of design. Good design in the modernist tradition must always pursue genuinely human(ist) interests; good design ultimately should have the end purpose of helping people accomplish their own purposes (Buchanan, 2005). Good design is, therefore, *instrumental* for satisfying human needs.

Although modernist normative ideals about design have been undermined by postmodernist critique, some modernist principles continue to be alive and well within contemporary design approaches[4]. The (arguably functionalist) argument that products exist ultimately to help human beings pursue their goals and, therefore, usability[5] ought to be privileged above cost, durability or aesthetics (Norman, 1990) is central for Human-Centred Design (HCD)—also known as User-Centered Design (UCD)[6]. As the name implies, HCD is an approach to designing that always starts by understanding the needs of the people using a product,

4 Related to the title of this paper, see Dieter Ram's highly influential principles for good design (Garreta Domingo, 2020).

5 Usability is understood here in terms of the standard definition, as the "extent to which a system, product or service can be used by specified users to achieve defined goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO EN 9241-11, 2018).

6 Looking closely, there are differences between HCD and UCD. While both are concerned with human needs, UCD may be regarded as a more compact subset of HCD as it categorises people as "users" from the start. For the sake of simplicity, in this paper, the terms are treated interchangeably.

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114 https://doi.org/10.34632/jsta.2020.9490

ideally not only at the individual level, but also at the collective level. HCD is the dominant paradigm in most contemporary design methodologies, particularly for Interaction Design (IxD) and User Experience (UX).

HCD's origins can be traced back to the early 1980s, to the multidisciplinary "Project on Human-Machine Interaction" from the Institute for Cognitive Science at the University of California, San Diego, headed by Don Norman. The findings of the project were condensed in the influential book, *User Centered System Design*[7] (Norman & Draper, 1986). Several of the ideas developed by Norman's group echoed the guidelines proposed by Gould & Lewis (1985) in the article "Designing for usability", which argued in favour of adopting an empirical approach in systems design based on thorough user research and intensive cycles of prototyping and testing.

7 The name "User Centered System Design" was originally an alliteration of the abbreviated name of the University of California, San Diego (UCSD). (Norman & Draper, 1986 iX) credit Paul Smolensky with having come up with the idea.

Nowadays, the main elements that characterise HCD methodologies include: analysing user practices empirically; employing interdisciplinary research; combining qualitative and quantitative data gathering and analysis; carrying out iterative prototyping and testing at various levels of fidelity; regarding the user as an expert possessing crucial (tacit) practical knowledge; and involving different stakeholders throughout the design process (Göransdotter & Redström, 2018). HCD methodologies are far from homogeneous and are also continuously evolving. However, as noted above, they have in common a commitment to the idea that technologies—particularly those involving computational tools and, therefore, smart automation—ultimately should be designed to improve people's lives. What such improvement means depends on the particular technology, the context where it is implemented, and the tasks it replaces.

Although HCD methodologies are now applied in most design fields, they have been traditionally employed in the design of computational or software-based devices—or at least, of their user interfaces (UI). The reason is that HCD emerged more or less around the time software-based tools started becoming consumer products. These devices were, and continue to be, necessarily complex[8], as we will see in the following section. Proponents of HCD saw design playing a crucial role in bridging the divide between these devices and their users. According to HCD, the task of the designer is to make the complexity of computational systems understandable for the user; meaning the designer would have to assume a mediating and epistemic duty.

8 Complexity is understood here as the quality of something (e.g., a system) consisting of many interconnected parts and whose behaviour requires considerable amounts of information to be described (Bar-Yam, 1997).

4. COMPLEXITY AND USER-CENTERED DESIGN

Every technological system has "an inherent amount of irreducible complexity" (Norman, 2010, p. 46). Complexity is present in—and arguably necessary for—many aspects of human life; this makes it unavoidable. Complexity is not problematic per se; it only becomes an issue when we do not understand it and, consequently, feel confused. Poorly understood complexity is, in this sense, complicated. Good design,

https://doi.org/10.34632/jsta.2020.9490

according to HCD, is not about making things less complex (i.e., simple), but less *complicated*, meaning, understandable (Norman, 2010).

Software-based devices are comparatively more difficult to use than their analogue counterparts. A word processor is significantly more complicated than any typewriter ever was. Software-based devices are also more challenging to use because they have a broader range of context-dependent states. A hammer, for example, does not have modifier keys to alter its behaviour; whereas gestures in a touchscreen and commands in a keyboard can trigger many different actions.

HCD argues that designers should seek a proper balance between the complexity of the underlying structure, behaviour, and limitations of a system, and the ways users intuitively conceive it. That means achieving a balance between the "implementation model" (i.e., the actual system's logic) and the "user's model" through a less complicated, usable "represented model" of the system (Cooper et al., 1995/2014) materialised in an adequately designed UI. Nonetheless, when a designer achieves a less complicated interaction for the user, the underlying complexity of the system increases accordingly. This paradox is known as "Tesler's law of the conservation of complexity" (Norman, 2010; see also Saffer, 2010, p. 136). It follows that when the UI is usable, the backend is likely complex; conversely, when the backend is optimised for the system's benefit, users will likely have to deal with a confusing implementation model.

Inherent in this oversimplified summary of HCD is the idea that designing involves mediating the relationship between humans and technologies; technologies whose goal is ultimately to serve human purposes; technologies which, paradoxically, are implicitly characterised as neutral instruments. Good design for HCD came to mean design that is above all *useful* (i.e., functional in achieving a given task), *usable* (i.e., understandable), and *desirable*. HCD's belief that a product that is designed to meet the genuine needs (i.e., needs that are supported by empirical evidence) of users is necessarily good betrays a modernist conception of human–technology relations.

Most design approaches based on HCD such as User Experience do consider the phenomenological dimension of technological interaction by focusing not just on usability but on the quality and enjoyment that products bring for users (Hassenzahl, 2010; Norman, 1990/2013). Nonetheless, their analysis of the *mediating* role of technologies and the implications it has across all aspects of human life remains limited.

5. HOW MEDIATION THEORY CAN HELP

Designing digital objects involves different processes than designing *analogue* or mechanical equivalents. Digitalisation has changed the traditional role of designers as "sculptural shape-makers" (Sudjic, 2015). Designers now have to explore ways in which objects can actively communicate their purpose and meaning. As noted earlier, HCD is mostly concerned with digital experiences, hence the prevalence of this

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114 https://doi.org/10.34632/jsta.2020.9490

approach in Interaction Design (IxD). As a multidisciplinary field where psychology, engineering, design, and aesthetics meet, IxD focuses on the ways people interact with software-based devices (Norman, 1990/2013). Ideally, interaction designers should not regard the product (just) as a concrete "thing" but as a "locus" of action and activity (Buchanan, 2001). IxD is therefore mainly concerned with shaping not just form and function, but mainly artificial behaviour. This behaviour manifests primarily as sequences of images on flat screens—increasingly supported by voice, sounds and haptics—through which the user navigates in a seemingly logical fashion. In summary, the core idea upon which IxD is grounded as a field is that what is designed is not things but *interactions* between humans and technologies.

The problem is that, as a concept, "interaction" —i.e., action in-between—is not enough for understanding technological artefacts and the complex relations human establish with and through them. Furthermore, as Verbeek (2015) argues, the very concept of interaction presupposes the separation of human subjects on one side and technological objects on the other. For contemporary philosophy of technology and mediation theory, in particular, such dichotomy is problematic because it overlooks that, through technologies, humans fabricate their circumstances, and therefore, themselves (see Hernández-Ramírez, 2019, pp. 20–21). The subject vs object dichotomy negates the artificiality of being human, the intrinsically human origin of technologies, and hence that, from an ontological standpoint, humans and their devices are mutually constitutive.

Verbeek argues that the specific relation between a human and a given technology is but a part of a broader relation between humans and their world "in which technologies play a mediating role". It follows that designers create not just interactions but rather human–world relations "in which practices and experiences take shape" (2015, p. 28). As previously noted, some HCD approaches do recognise the broader phenomenological implications of design and hence focus not just on interactions but on the "user experience" elicited by a product. However, for UX practitioners, human–technology relations continue to be defined in terms of functionality and use: a pleasurable and therefore *good* experience is ultimately dependent on the artefact being thoroughly functional (i.e., because it satisfies a specific human need). The problem with this view, as Verbeek notes, is that many relations we have with contemporary technologies cannot be adequately characterised solely in terms of *use* and *goals*—e.g., when dealing with recommendation algorithms it is not entirely clear who the *user* actually is. Moreover, goals do not exist independently from the technologies that enable them, a prime example is social media, which has brought "new types and dimensions of social relations" not intended initially when these systems were designed (2015, p. 28). Whenever a given technology becomes part of a sociotechnical system, unexpected consequences are bound to emerge.

According to mediation theory, Human–technology relations are more complex than the instrumentalist view that HCD inherited from

https://doi.org/10.34632/jsta.2020.9490

modernism might concede. Technologies are not just extensions of human capacities; they can also enhance, augment or completely replace them. Technologies are not opposed to human nature but rather are intrinsically bound to it. Characterising technologies foremost as *mediators* allows us to inspect how they shape our understanding of the world and ourselves. Seen like this, designing interactions implies designing a system that will mediate the relation between humans and their environment and thus, implies designing how they live their lives (Verbeek, 2015).

Designers are by no means morally neutral agents (Buchanan, 2005). They inscribe values, preferences, and prejudices about the world in their designs. Their choices are embedded in the *solutions* they find to address the *problems* they encounter. Since each product is intended to be used by another human being and that usage is directed towards a given goal, it is not difficult to concede that every product embodies an argument about how we should carry out that activity, and, therefore, about how we should lead our lives (Buchanan, 2001). Hence, designing interactions implies designing not only the technological objects but also the human subject (i.e., the *user*) who will interact with it. Bluntly put, designing technologies is, in a way, designing human beings (2015, p. 28). Granting the previous arguments, we can agree that "designing is a form of ethics" (Buchanan, 2005, p. 508) or at least a way of "doing ethics by other means" (Verbeek, 2006). A good approach to designing hence should involve taking a position about the role that artefacts play in human lives and anticipating the consequences they might bring for human welfare.

6. TOWARDS AN ETHICALLY MINDED HCD

HCD is the dominant paradigm in Interaction Design and User Experience design. From a purely methodological standpoint, HCD should be one of the best means to minimise poorly implemented automation. However, our world continues to be further populated by systems that incite users with "perverse incentives" (Loh & Misselhorn, 2018), dark patterns (Monteiro, 2019), and algorithms that deepen structural inequalities. Smart automated systems are already complex networks of algorithms, sensors, human agents interacting under changing contexts and at different time scales (Woods, 2016). The ironies of automation discussed in section two are further increased by the risks and complexities of contemporary computational infrastructure, namely, the fact that even the simplest interconnected system is embedded within a growing ecosystem of "balkanised operating systems, stacks of numerous protocols, versions, frameworks, and other packages of reusable code" (Girardin, 2019). Designers who follow HCD should not just focus on making the complexity of these systems understandable. Given how many aspects of people's well-being now depend on them, HCD design should also be about protecting people from the accidental and deliberate misuses of these technologies. Designers endorsing HCD can no longer be mere facilitators

https://doi.org/10.34632/jsta.2020.9490

of usable technologies but also, or rather, mostly, "gatekeepers" (Monteiro, 2019) for their users. HCD needs more than *empathy* towards users; it needs to assume a clear stance about the role of design on technological development and, therefore over human life, since products have short and long-term consequences for individuals, society, and the natural environment at large (Buchanan, 2005). When looking at products as mediators of human experience, it becomes challenging to distinguish design from ethics and politics.

HCD needs to rethink how it characterises good design. The idea that thoroughly understanding users' needs with *just enough science* and coming up with products that are usable and fully satisfy those needs is sufficient for design to be good has to be challenged. Reducing applied ethics in design to the correct application of a methodology is a rather poor approach, even from a normative moral standard. Furthermore, an approach that focuses solely on satisfying needs, regardless of how genuine the empirical evidence shows them to be, assumes that those needs exist *independently* of the technology that will satisfy them. Such instrumentalist view perpetuates the notions that technologies are neutral, and that subjects and objects can *be* independently of each other.

An argument can be made that unethical design can be curtailed through regulation. Designers, being experienced professionals, could simply abide to the moral standards of fairness, honesty, and loyalty expected in any business relationship (Buchanan, 2005, p. 505). As moral agents, designers are expected to follow the standards necessary for maintaining product integrity, such as complying with laws and regulations governing the safety and reliability of products. After all, these regulations exist for a reason. The problem with this argument, however, is that regulation tends to lag rather than accompany technological developments.

As shown by the recent scandal involving Joichi Ito and the MIT Media Lab, academic AI ethics has been skilfully manipulated into becoming mere whitewashing for otherwise unethical forms of automation (Hao, 2019; Ochigame, 2019) mostly because this approach has grown around the idea of voluntary compliance. Furthermore, designing an embedded Asimov-style moral code into smart automated systems is a considerably difficult task since it involves reducing the complexities of ethical principles to procedural steps (Ceglowski, 2016). Academic committees counselling on potential ethical mishaps with no actual influence over design processes is a recipe for inaction. Then, it should be up to designers to take the lead.

The design process itself should be imbued with an ethical framework. The responsibility should be put in the designer, as well as in the product of her work. Ethics cannot be regarded as moral aspirin but as an integral aspect of doing design. This ethical framework might involve adopting a starker and more careful attitude towards automation, it should include following a kind of *via negativa*[9] approach that focuses not on what could be beneficial, but on everything that could go wrong should a given form of automation is allowed to come into fruition. Nevertheless, for this

9 That is, "the principle that we know what is wrong with more clarity than what is right, and that knowledge grows by subtraction" (Taleb, 2018, Book I); see also Taleb (2012, Book VI) for a more detailed description.

https://doi.org/10.34632/jsta.2020.9490

to happen, designers need to see beyond instrumentalism and incorporate a critical framework for thinking about human–technology relations and how artefacts influence our multileveled engagement with the world.

An ethically minded HCD approach should start by asking what the purpose of a given technological system is, e.g., asking whether its implementation will improve or negatively affect human life. The former implies considering whether a given process should be automated at all. Automation, as we have seen, is an attempt to reduce the complexities of a given circumstance into piece-meal deterministic steps. However, because many aspects of any given process cannot be automated there is always room for unknown complications to emerge. These complications are further increased because, the training of smart automation is always done with available information, this narrowing exponentially increments the space of unknowns when dealing with complex situations, as the problem of induction illustrates. Consequently, what I propose here is that rather than focusing on the *how's* of smart automation, an ethically-minded HCD should always begin with the *why's*. That is, the designer should ask what kind of trade-offs should be assumed if her design is indeed implemented.

As was earlier noted, a goal of HCD is making technologies understandable. For an ethically minded HCD, this should imply making systems that are more transparent and honest about their limitations, potential side effects, and dependencies. This implies the system should be obtrusive under certain circumstances and more honest about the *frictions* it might bring to its users. An ethically minded HCD should always analyse user tasks and goals within a broader context. Technologies are never neutral or inert, their introduction *always* affects a given context; their being and meaning are always situated, and the very fact of using a given instrument continually modifies the task itself. An ethically minded HCD should be keenly aware of the mediation role of automated systems and technologies at large. This implies a broader understanding of *experience* that strives to eliminate the barrier between aesthetics and ethics. Understanding automated systems as situated also means adopting a sociotechnical systems approach to design. An ethically minded HCD requires awareness of designer's blind spots about technology, that is, about the hidden assumptions they have about what a given system can and should do for users.

What is needed then is not just integrating some generic concepts about ethics into HCD approaches but making applied design ethics a core aspect of HCD. Mediation theory and other approaches to philosophy of technology, as well as philosophy of design can be instrumental in this regard, helping designers to develop skills to question basic assumptions about their design processes. In other words, designers should be equipped with the tools to approach the philosophical dimension of design issues and their assumptions about their beliefs and practices, treating them as problem spaces open to reasonable discussion. HCD needs to reconsider its standard for good design and therefore steer way from instrumentalism. These changes, however, cannot happen overnight; for

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114 https://doi.org/10.34632/jsta.2020.9490

them to come, professional designers and academics must take action. An excellent place to start is design education. Young designers should be trained not only in design research but also have a critical, situated, understanding of technology and the tools to *problematise* it from a design standpoint. Design education should thus be supported by professional applied ethicists with sufficient knowledge of design processes and practices, and curricula should be updated to include a philosophical tool to support design research.

7. CONCLUDING REMARKS AND FUTURE WORK

This paper began with the assumption that lousy automation is "bad design". It argued that good design should not just make complexity understandable. As we have seen then, it is virtually impossible to build error-free *traditional* automated systems because humans are integral to their operation. Good design is not just empathy; good design is research, iteration, questioning, *tire-kicking* and diversity. This paper has shown there is a lack of correspondence between the ideals of HCD and what in fact, comes out of design processes. Based on this critique, this paper argued that HCD should provide a richer, fairer, ethically cantered, framework for designers. However, to do so, HCD needs to make explicit its stance on technologies, recognising their mediating role and thus to critically look at the moral, political and social consequences of digital devices.

HCD should recognise that design is political. This realisation should be the starting point to determine what is good design and how that measure will affect design in the years to come. This paper, however, clearly has *not solved* HCD once and for all, nor has it proven beyond doubt what good design should mean nowadays. There is still much work to do in this regard. However, it has managed to start calling into question the neutrality that HCD practitioners often assume this approach has. This paper regards design as a collective enterprise that requires designers to work not on behalf of people but *with* people; that asks of designers to behave as gatekeepers with *skin in the game* that strive to achieve not minimum viable prototypes, but minimum ethical ones. Ultimately, designers should ask what purposes technologies have in human life. They should ask themselves what our human project for the future is. Future work on this regard should include a thorough analysis of the hidden assumptions that design methodologies have about human–technology relations, but also developing new frameworks. The arguments developed in this paper can serve as pointers for doing so.

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114 https://doi.org/10.34632/jsta.2020.9490

REFERENCES

Autor, D. (2014, September). *Polanyi's paradox and the shape of employment growth*.
http://dx.doi.org/10.3386/w20485

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775–779
https://doi.org/10.1016/0005-1098(83)90046-8

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Addison Wesley.

Buchanan, R. (2001). Design and the new rhetoric: Productive arts in the philosophy of culture. *Philosophy and Rhetoric, 34*(3), 183–206.
https://doi.org/10.1353/par.2001.0012

Buchanan, R. (2005). Design ethics. In C. Mitcham (Ed.), *Encyclopedia of science, technology, and ethics: Vols 2 (D–K)* (pp. 504–510). Macmillan Reference.

Ceglowski, M. (2016, October 29). Superintelligence: The idea that eats smart people. *Idle Worlds*.
http://idlewords.com/talks/superintelligence.htm

Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About face: The essentials of interaction design* (4th ed.). John Wiley & Sons.

Danaher, J. (2018). Toward an ethics of AI assistants: An initial framework. *Philosophy & Technology, 31*(4), 629–653.
https://doi.org/10.1007/s13347-018-0317-3

Flusser, V. (2014). *Gestures*. University of Minnesota Press.

Garreta Domingo, M. (2020). Dieter Rams: 10 Timeless Commandments for Good Design. *Interaction Design Foundation.*
https://www.interaction-design.org/literature/article/dieter-rams-10-timeless-commandments-for-good-design

Girardin, F. (2019, January 16). *When automation bites back.*
http://blog.nearfuturelaboratory.com/2019/01/16/when-automation-bites-back/

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM, 28*(3), 300–311.
https://doi.org/10.1145/3166.3170

Göransdotter, M., & Redström, J. (2018). Design methods and critical historiography: An example from Swedish user-centered design. *Design Issues, 34*(2), 20–30.
https://doi.org/10.1162/desi_a_00483

Graeber, D. (2018). *Bullshit jobs: A theory.* Simon & Schuster.

Hao, K. (2019, December 27). In 2020, let's stop AI ethics-washing and actually do something. *MIT Technology Review*.
https://www.technologyreview.com/s/614992/ai-ethics-washing-time-to-act/

Journal of Science and Technology of the Arts, Vol. 12, n. 3 (2020): pp. 100-114 https://doi.org/10.34632/jsta.2020.9490

Hassenzahl, M. (2010). *Experience design: Technology for all the right reasons* (S. L. on Human-Centered Informatics, Trans.; Vol. 3, pp. 1–95). Morgan & Claypool Publishers LLC.
https://doi.org/10.2200/s00261ed1v01y201003hci008

Hernández-Ramírez, R. (2019). On false augmented agency and what surveillance capitalism and user-centered design have to do with it. *Journal of Science and Technology of the Arts, 11*(2), 18–27.
https://doi.org/10.7559/citarj.v11i2.667

ISO EN 9241-11. (2018). *Ergonomics of human-system interaction—part 11: Usability: Definitions and concepts. International Organization for Standardization*.
https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. *Minds and Machines, 27*(4), 575–590.
https://doi.org/10.1007/s11023-017-9417-6

Loh, W., & Misselhorn, C. (2018). Autonomous driving and perverse incentives. *Philosophy & Technology, 32*(4), 575–590.
https://doi.org/10.1007/s13347-018-0322-6

Martinho-Truswell, A. (2018, February 13). To automate is human. *Aeon*
https://aeon.co/essays/the-offloading-ape-the-human-is-the-beast-that-automates

Monteiro, M. (2019). *Ruined by design: How designers destroyed the world, and what we can do to fix it.* Mule Design.

Norman, D. A. (2013). *The design of everyday things* (Revised and expanded edition). Basic Books.

Norman, D. A. (1990). *The Psychology of Everyday Things.* Perseus Books Group.

Norman, D. A. (2010). *Living with Complexity*. MIT Press.

Norman, D. A., & Draper, S. W. (Eds.). (1986). *User centered system design.* Lawrence Erlbaum Associates.
https://doi.org/10.1201/9780367807320

Ochigame, R. (2019, December 20). The Invention of "Ethical AI": How Big Tech Manipulates Academia to Avoid Regulation. *The Intercept.*
https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

Parsons, G. (2015). *The philosophy of design.* Polity Press.

Saffer, D. (2010). *Designing for interaction: Creating innovative applications and devices* (2nd ed.). New Riders.

Strauch, B. (2018). Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems, 48*(5), 419–433.
https://doi.org/10.1109/thms.2017.2732506

Sudjic, D. (2015). *B is for Bauhaus, Y is for YouTube: Designing the Modern World from A to Z*. Rizzoli Publications.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder.* Random House.

Taleb, N. N. (2018). *Skin in the game: Hidden asymmetries in daily life*. Random House.

Verbeek, P.-P. (2015). COVER story: Beyond interaction. *Interactions, 22*(3), 26–31.
https://doi.org/10.1145/2751314

Verbeek, P.-P. (2006). Materializing morality. Science, T*echnology, & Human Values, 31*(3), 361–380.
https://doi.org/10.1177/0162243905285847

Vermaas, P., Kroes, P., van de Poel, I., Franssen, M., & Houkes, W. (2011). *A philosophy of technology: From technical artefacts to sociotechnical systems* (Vol. 6). Morgan & Claypool Publishers LLC.
https://doi.org/10.2200/s00321ed1v01y201012ets014

Woods, D. D. (2016). The risks of autonomy. *Journal of Cognitive Engineering and Decision Making, 10*(2), 131–133.
https://doi.org/10.1177/1555343416653562

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*(1), 75–89.
https://doi.org/10.1057/jit.2015.5

ACKNOWLEDGEMENTS

This study was supported by UNIDCOM under a Grant by the Fundação para a Ciência e Tecnologia (FCT) No. UIDB/DES/00711/2020 attributed to UNIDCOM/IADE – Unidade de Investigação em Design e Comunicação, Lisbon, Portugal. The author would like to thank the organizing committee of xCoAx for providing a most excellent forum to present his ideas.

Article received on 01/09/2020 and accepted on 11/11/2020.