AI & SOCIETY
https://doi.org/10.1007/s00146-020-01004-z

OPEN FORUM

A taxonomy of human–machine collaboration: capturing automation and technical autonomy

Monika Simmler[1] · **Ruth Frischknecht**[2]

Received: 15 January 2020 / Accepted: 10 June 2020

Abstract
Due to the ongoing advancements in technology, socio-technical collaboration has become increasingly prevalent. This poses challenges in terms of governance and accountability, as well as issues in various other fields. Therefore, it is crucial to familiarize decision-makers and researchers with the core of human–machine collaboration. This study introduces a taxonomy that enables identification of the very nature of human–machine interaction. A literature review has revealed that automation and technical autonomy are main parameters for describing and understanding such interaction. Both aspects must be carefully evaluated, as their increase has potentially far-reaching consequences. Hence, these two concepts comprise the taxonomy's axes. Five levels of automation and five levels of technical autonomy are introduced below, based on the assumption that both automation and autonomy are gradual. The levels of automation were developed from existing approaches; those of autonomy were carefully derived from a review of the literature. The taxonomy's use is also explained, as are its limitations and avenues for further research.

Keywords Human–machine collaboration · Taxonomy · Automation · Autonomy

1 Introduction

In the digital age, human–machine interaction is an essential part of everyday life, such as when pilots fly airplanes, physicians make use of diagnostic programs, workers operate industrial machines, and Uber drivers receive messages about jobs. All of these examples share a commonality: a task is accomplished in collaboration with an advanced machine. In many areas, actions are increasingly the result of such human–machine collaborations (Kirchkamp and Strobel 2019). Machines are taking over more and more tasks that previously were performed by humans (Vagia et al. 2016). However, these advances are not limited to the optimization of automation processes, and, therefore, not relegated to the simple delegation of functions and operations to technical artefacts. Rather, through the rise of artificial intelligence and machine learning, technology has obtained abilities that formerly were reserved solely for humans (Jordan and Mitchell 2015). Because of these two developments (that can also be referred to as automation and autonomization), human–machine collaboration has not only become more prevalent, but also more complex.

For describing human–machine interaction, the concept of socio-technical systems has been introduced. The term indicates that technical components and people are included as inherent parts of a system (Sommerville 2007). In its simplest form, a socio-technical system is composed of one human and one technical component. Of course, socio-technical constellations usually take on more complex forms involving many different technical agents and different people or social systems (e.g., organizations). However, every analysis of a socio-technical system implies the analysis of a human–machine collaboration which fulfils a certain function for a certain duration, regardless of its complexity. As the implementation of technology has social, ethical, legal, and economic implications, the setup of socio-technical systems calls for careful consideration (Martin 2018), requiring one to define the distribution of roles according to ability, authority, and control (Flemisch et al. 2012). The design of human–machine collaboration affects its governance

✉ Ruth Frischknecht
ruth.frischknecht@unisg.ch

[1] Law School, University of St. Gallen, Bodanstrasse 3, 9000 St. Gallen, Switzerland

[2] Institute for Systemic Management and Public Governance, University of St. Gallen, Dufourstrasse 40a, 9000 St. Gallen, Switzerland

Published online: 02 July 2020

(Danaher et al. 2017) and accountability (Shin and Park 2019), as both go hand-in-hand with the underlying distribution of agency (Beck 2015; Matthias 2004; Pagallo 2017). Thus, it shapes the accompanying monitoring measures, as well as the system's administration (Shin and Park 2019).

Due to the importance of the constitution of socio-technical systems and their underlying human–machine collaboration, these phenomena pose a challenge for management, the law and other disciplines. Organizations and managers must oversee possible changes that the implementation of a new technical system might bring (Shneiderman 2016) and understand its broader impacts (Nunes and Jannach 2017; Shin and Park 2019). However, decision makers are often unfamiliar with the fundamentals of implementing human–machine collaboration (Shin and Park 2019). Managing technology requires a full grasp of the nature of the human–machine collaboration. Comparably, legal assessment equally asks for capturing the distribution of agency, because it significantly affects the attribution of responsibility (Simmler 2019). The digital age confronts the law with a reality in which human–computer interaction is the rule rather than the exception (Wein 1992), challenging practitioners to deal with technology with which they are unfamiliar.

Understanding the core of a socio-technical collaboration is no longer a task solely for engineers, but also for decision makers and experts in different fields and disciplines. As illustrated above, not only answering questions regarding if, but also how to implement human–machine interaction within socio-technical systems are decisive managerial tasks. Above all, such implementation requires that those responsible understand and capture the essence of human–machine collaboration. Therefore, the goal of this research is to provide a comprehensive yet straightforward taxonomy for classifying human–machine collaboration. Taxonomies are valuable tools for understanding and analyzing complex phenomena (Nickerson et al. 2009). Determined by their aim (Lambe 2007; Nickerson et al. 2009), they store knowledge in a concise form and reduce complexity (Lambe 2007). The goal of the taxonomy presented here is to categorize human–machine collaborations by their most fundamental socially perceived characteristics. Decision makers and researchers will now be equipped with a tool that enables them to quickly grasp the implications of the human–machine collaboration at issue. A taxonomy that focuses on the social perception of socio-technical phenomena will help to evaluate such collaborations in terms of their consequences for different areas and with respect to a diversity of aspects.

Following this objective, this research first describes the method by which this taxonomy was developed. Next, existing taxonomies and classifications are introduced, leading to the elaboration of the present taxonomy's theoretical foundation. Hereafter, the taxonomy is developed and described in detail. In addition, the approach is further clarified with examples. Finally, the scope and limitations of the taxonomy and avenues for further research are discussed before conclusions are drawn.

2 Methodical framework

The approach to developing this taxonomy consisted of a literature review and consideration and elaboration of existing taxonomies and categorizations. We searched the most common databases using terms such as "technology," "classification," and "taxonomy." Whenever possible, we based our analysis on existing taxonomies and classification systems. However, because some of these did not fully meet our objectives, we extended and adapted what we found. Based on the results of this process, we followed three steps to develop the taxonomy presented here.

We first evaluated the core parameters and axes. In so doing, we evaluated whether each parameter captured the essence of human–machine interaction. A parameter was considered relevant if it significantly shaped the nature of the human–machine collaboration and clearly displayed differences between the human and technical components.

In the second step, we derived the parameter levels. We mapped each level on a continuum ranging from "none" to "all." We introduced new levels when there were significant shifts in social and technical differences (i.e., when a technical component was assigned a new task or capability). To determine these levels, we focused especially on their (social) implications (i.e., their impact on social perceptions and the attribution of responsibility) and their influence on accountability.

In a third step, we validated the parameters and levels with regards to their clarity, comprehensibility, and practicability. In this step, we wanted to make sure that the taxonomy would be useful to people with little or no technical expertise. The goal was to present an instrument for capturing the basic characteristics of interdisciplinary relevance of human–machine collaboration.

3 Theoretical background

3.1 Conceptualizing socio-technical phenomena

In the past several decades, there has been a multitude of approaches to capturing human–machine collaboration. The long-standing tradition has focused on different levels of automation (Vagia et al. 2016). These taxonomies describe different role allocations (Kaber 2018) and offer descriptions of what tasks the "human operator" and "computer" are

assigned within a collaboration. Usually, the levels of automation range from fully manual to fully automated (Vagia et al. 2016). Fully manual describes a situation in which the human is fully in charge. Conversely, fully automated indicates that the human operator is completely out of the loop and, therefore, is obsolete (Parasuraman et al. 2000). Such taxonomies center on the interaction between human and machine, the level of which is derived from variations in task allocation (Endsley 1987; Sheridan and Verplank 1978; Weyer 2006). However, there seems to be less consensus about what and how many levels lie in between these points. Some approaches describe four (Endsley 1987), others ten (Sheridan and Verplank 1978), and some even more (Riley 1989). A recent study described different levels of automation with regards to autonomous driving (NHTSA 2013). There, the scale ranged from "no automation" to "autonomous", implying that autonomy is the highest form of automation and reflecting a slightly different stance. Solely focusing on role allocation, however, is not the only way of describing human–machine interactions. There are models that also take into account the process by which a task is completed. By categorizing every stage of a given process or decision, the level of automation can be distinguished even more precisely, depending on what stage of the process is being automated and to what extent (Parasuraman et al. 2000; Proud et al. 2003; Weyer 2006).

A different research stream centers on a technology's abilities. One line of thought has focused on naming and discussing the capabilities of an advanced technology, without further mapping. In this debate, scholars have discussed conditions for technical autonomy and agency. The capacity to react to changes in the environment (Alonso and Mondragón 2004; Franklin and Graesser 1997) and the general ability to interact (Floridi and Sanders 2004; Misselhorn 2015) have both been proposed. Adaptability has been repeatedly identified as relevant, proposed as either an optional (Franklin and Graesser 1997) or necessary (Floridi and Sanders 2004; Misselhorn 2015; Russell and Norvig 2014) condition for technical agency. An autonomous agent has been described as having the ability to perceive its environment and then learn from and adapt to it (Alonso and Mondragón 2004; Pagallo 2017; Santosuosso and Bottalico 2017; Sartor and Omicini 2016) or to sense, plan and intentionally act upon an environment without external control (Beer et al. 2014). A study investigating what consumers regard as "intelligent" when it comes to technology came to similar conclusions. Product intelligence is said to be composed of the key dimensions of autonomy: an ability to learn, reactivity, an ability to cooperate, humanlike interaction, and personality (Rijsdijk et al. 2007).

While these approaches suggest preconditions and basic characteristics, none conceptualize connectivity among the parameters or clearly topologize their impact on autonomy. Yet with regards to technical agency, there have been a number of approaches suggesting different stages or levels. According to these approaches, technical agency is not a binary category, but rather varies in terms of degree. For example, it was suggested that one must distinguish different activity levels when dealing with technical agency, using a scale ranging from "passive" (describing artefacts operating as mere tools) to "transactive" (indicating actions based on intelligent associations stemming from self- and external reflection) (Rammert 2009). Advanced technology is assumed to challenge the passive nature of technology, blurring the boundary between technical functionality and human agency. Rammert and Schulz-Schaeffer (2002) identified three levels of agency: causality as the capacity to cause change, contingency as the ability to recognize action alternatives, and the highest level, intentionality, which is defined as intentional reasoning engaging in action. Thürmel (2015) added another gradualization by proposing that the degree of agency is constituted by four dimensions: activity and adaptivity for individual agency as well as interaction and personification of others for joint agency.

There have also been other approaches to combine different variables into one categorization of technical systems, such as autonomy, field of application, and morphology (Onnasch et al. 2016) or automation and complexity (Janssen and Kuk 2016), the latter following the assumption that the more complex a technical system is the opaquer it gets. While these approaches mainly center on the technical side, the typology provided by Gransche et al. (2014) additionally describes the human role in human-technology interaction, while comparing and linking technical and human autonomy with control. According to the authors, the "normative" type of autonomy and control is reserved only for humans; technical systems achieve "strategic" and "operative" autonomy and control (Gransche et al. 2014).

3.2 Synthesis

To sum up, there are different taxonomies for describing human–machine collaboration, and these follow different aims. However, the literature review has shown that there are two basic questions underlying these existing frameworks. One stream describes different role allocations between humans and technical systems and asks who does what? (Kaber 2018). This is the core question supporting the concept of automation. Taxonomies centering on automation usually address information systems experts and their pursuit of technical optimization. Such taxonomies are accurate in their capture of basic design issues. Nonetheless, task allocation is not the only point to consider when implementing and optimizing human–machine collaboration. The second stream takes technical functions and properties into account more directly. They focus on intelligence (Rijsdijk et al.

2007), autonomy (Beer et al. 2014), or agency (Thürmel 2015), and, therefore, on the technical system's abilities. These approaches are strongly interwoven, and different scholars have reached similar conclusions indicating that the primary concern is to describe a technology's independence. The basic question behind these approaches is: what degree of freedom does the technology have when completing the assigned task? Therefore, they center on technical autonomy.

In sum, a first analysis revealed that the underlying most fundamental characteristics of human–machine collaboration are automation and autonomy. Current approaches to understanding and conceptualizing such socio-technical interaction focus either on automation or technical autonomy (or comparable parameters). Sometimes, these notions are used interchangeably (Vagia et al. 2016) or mapped onto the same continuum (NHTSA 2013). We suggest that the allocation of situational executive control (automation) and independence of the technical component assigned (autonomy) only in combination comprise the core parameters of human–machine collaboration.

4 Taxonomy

4.1 The axes: automation and autonomy

Although there is no universal definition, it is generally agreed that automation refers to a state in which a technical system performs a formerly human task or parts of that task, respectively (Parasuraman et al. 2000; Vagia et al. 2016). In fulfilling this task, the machine runs without a human operator (Hertzberg 2015; Nof 2009). As explained above, the task assigned to the technical system might vary in its scope, i.e., be partially or fully automated (Vagia et al. 2016). With regards to human–machine-collaboration, the level of automation describes the extent of the machine's contribution to the joint performance. It is crucial to determine whether and to what extent humans are still "in the loop" with regards to acting and decision-making (Santosuosso and Bottalico 2017; Sartor and Omicini 2016). Designing human–machine collaboration requires the careful determination of the tasks to be automated, and to what extent (Parasuraman et al. 2000). Thoughtful design decisions depend upon a thorough understanding of such task allocation (Vagia et al. 2016).

While an appropriate definition of the level of automation is crucial for capturing this allocation of control, it neglects the nature of advanced technologies. Advanced machines can be more or less independent from their human creators and operators (Rammert and Schulz-Schaeffer 2002). Technical autonomy addresses this notion. Traditionally, autonomy is understood as self-governance, self-sufficiency, or self-directedness (Bradshaw et al. 2004; Vagia et al. 2016); it is a relational concept describing the degree of independence from something, such as the specific influence of another entity, the environment, or internal restraints (Castelfranchi and Falcone 2004; Müller-Hengstenberg and Kirn 2016; Verhagen 2004). Usually, the possibility of independently developing action alternatives and acting according to one's own preferences serve as prerequisites for autonomous behavior (e.g., Müller-Hengstenberg and Kirn 2016). In a Kantian view, autonomy requires the ability to adopt maxims to govern one's action and involves the power to make laws for oneself (Hilgendorf 2017; Korsgaard 2014; Misselhorn 2015). This association with Kant's moral philosophy requires a careful use of the term; however, its use in describing technical agents is far less demanding (Misselhorn 2015). As mentioned above, the debate surrounding the autonomy of advanced technology is closely related to the discussion of the possibility of and conditions surrounding technical agency per se, as these are often regarded as mutually conditional (Misselhorn 2015). Autonomy is, therefore, closely linked to the concept of agency.

How autonomously a technical component operates is important to understand, because this affects not only the limits of human control, but, furthermore, also other variables such as the explainability, traceability and, predictability of the technical component's action (Balkin 2015). Thus, for an initial assessment, especially in terms of consequences, it is important to determine how the task is fulfilled by the technical system. Therefore, looking at technical autonomy is equally pivotal. Evaluating a human–machine collaboration demands the determination of both values: the extent to which the task is automated (automation) and how autonomous the technical system operates when pursuing that automated task (autonomy). We conceptualize both automation and autonomy as gradual, indicating that these are not binary categories but rather vary in degree. This suggests a taxonomy composed of two axes that consist of different levels.

This taxonomy focuses solely on technical autonomy. Conversely, the autonomy of the human as a part of this human–machine collaboration is generally regarded as given and more substantial than what technology can ever reach (Misselhorn 2015; Rammert 2009). Deliberately focusing on technical autonomy, this taxonomy does not map or describe the concept of human autonomy in detail. A further division of human autonomy would not add value to the understanding of human–machine collaboration. Although the taxonomy focuses on technical autonomy, its description still includes a human perspective: the extent to which the actions of technology are comprehensible to human counterparts. Thus, the levels of technical autonomy include the human perspective within the human–machine collaboration, though not directly.

The following examples highlight the difference between automation and autonomy and clarify the benefit of including these two axes in one taxonomy. A physician and diagnostic program together collaborate when they pursue the task of diagnosing patients. Fulfillment of this task can be automated to different degrees (i.e., a computer program can be assigned a varying part of the task). It can, for example, suggest diagnoses when certain symptoms are entered or the program can reach a single diagnosis and even order medication without human influence. In addition, the technical component can be more or less autonomous in a socio-technical constellation. For example, the diagnostic system can only be a very cognitively limited computer program with a restricted dataset and clearly defined variables and algorithms. If it is more autonomous, it is an open system that obtains data from networks or other technical agents. Furthermore, it can learn from diagnoses already conducted and adapt its behavior accordingly. These differences relate to its level of autonomy.

A self-driving car will serve as a second example. Here, the task of operating the car is fully automated. No human intervention is needed. Therefore, a self-driving car reaches the highest stage of automation. In terms of autonomy, however, the evaluation may vary from car to car. There are self-driving cars whose operation is fully determined, and a given body of information always leads to the same predefined output. Due to the absence of machine learning, the car does not alter its behavior. Yet there may also be cars that reach notably higher levels of autonomy. Some self-driving cars not only learn and adapt, but are also connected to other agents, resulting in an open, multi-agent system. Consequently, even within the scope of fully automated driving, there is a range of different types of human–machine interplay, as the autonomy of the technical aspect can vary significantly.

To again underscore the important distinction between the two axes and the value they add to accurately capturing socio-technical systems, it is useful to revisit what lies at the taxonomy's core. It centers on human–machine collaboration of a certain permanence and distinctiveness, consisting of one or multiple human actors and one ore multiple technical agents (Weyer and Reineke 2005). If one focuses only on a single task taken over by a machine, that task can be described as embodying full automation. However, conceptualizing socio-technical phenomena means focusing on a human–machine collaboration not only executing a single act, but also fulfilling a distinct and permanent function. During this persisting collaboration, the system as a whole can be more or less automated, and within these automated tasks, the machine can be more or less autonomous. Both axes define the human–machine interaction at its core.

4.2 Levels of automation

We draw our approach to defining the levels of automation from Endsley's (1987) four levels, because this study is concise and comprehensible and does not neglect the core aspects of task allocation in the context of human–machine collaboration. We complemented Endsley's levels of automation with an additional level borrowed from Weyer (2006). Different from most taxonomies (Vagia et al. 2016), we did not include a level for manual (and, therefore, solely human) operations, because our taxonomy was designed only to classify human–machine collaboration. This implies that there is already a minimum amount of technical participation, which renders a level for classifying manual tasks unnecessary. The five stages are shown in Table 1 and described in detail thereafter.

4.2.1 Level 1: Offers decisions

Our taxonomy begins with what can be called decision-support systems (i.e., constellations in which the technical component makes suggestions). The machine, usually a computer program, makes recommendations to the operator. The operator chooses whether to act accordingly (Endsley 1987). In this early stage of automation, the system presents options; however, it is still the human operator who selects and, thus, decides.

Table 1 Levels of automation with their main features

Level	Description	Explanation
1	Offers decision	Technical component suggests options and the human decides
2	Executes with human approval	Technical component acts after human approves
3	Executes if no human vetoes	Technical component acts unless human vetoes
4	Executes and then informs	Technical component acts independently and human is informed about the actions carried out
5	Executes fully automated	Technical component carries out actions independently without informing human

4.2.2 Level 2: Executes with human approval

In the second level of automation, the machine executes with human approval. Here, the technical component makes a recommendation that it carries out if the operator concurs (Endsley 1987). This marks an increase in automation in that the system selects among options and presents only the "best" alternative. The selection process is no longer carried out by the human operator, but rather by the technical component. However, at this level it is still the human who makes the final decision, either by approving or rejecting the system's proposition.

4.2.3 Level 3: Executes if no human vetoes

In our third level of automation, the technical system executes if no human veto occurs. According to Endsley (1987), this means that the system recommends a single option, which it will carry out unless the operator vetoes that decision. This represents an increase in automation, because the technical system selects an option and automatically decides. However, the human operator remains a part of the process, because the operator can correct the action through its veto power. The decisions made by the system still require human consent. However, this no longer happens in advance, only after the component has already reached a decision. This influences the social perception of the operator's role.

4.2.4 Level 4: Executes and then informs

Further increasing the automation, at this level the technical system executes and then informs. We borrowed this level from Weyer (2006) and use it to augment Endsley's (1987) approach. The system selects an option and then automatically acts accordingly. However, the human operator still has a role in the process, because the system informs the human when it acts. Even if a human operator no longer plays an active role, they remain informed about the system's actions. The knowledge of these actions is relevant for social perception. Therefore, it differs from Level 5, where no human is made aware of the automated process.

4.2.5 Level 5: Executes fully automated

At the fifth level, the technical system executes fully automated. This means that the human is neither informed about nor actively participates in the action. The technical system is fully in charge, rendering human action obsolete.

4.3 Levels of autonomy

There was no classification of technical autonomy available that met with the scope of our taxonomy (i.e., that described the independence of the technical component while completing an assigned task). Therefore, we could not simply borrow different levels of autonomy as we did when typologizing automation. After reviewing the literature, we isolated the four dimensions we considered most crucial when determining the different levels of autonomy.

We tend to call something "autonomous" if it appears untraceable. Technical autonomy is generally said to describe systems that to a certain degree are independent and not fully determined (Verhagen 2004; Castelfranchi and Falcone 2004; Müller-Hengstenberg and Kirn 2016). Thus, technical autonomy involves a certain lack of traceability and certainty regarding the system's performance, and concerns both, transparency and determinism. Conversely, the absence of technical autonomy means full determination and traceability. A technical system that meets this condition is called deterministic. In a deterministic system, input A always equally leads to output B, and all execution steps in between are specified and transparent (Loh and Loh 2017). A technical system is considered determined when every possible condition of the system unambiguously results in a subsequent condition of $n+1$ (Müller-Hengstenberg and Kirn 2016). However, such a determined system does not necessarily need to be transparent in all of its execution steps, leaving the observer or user ignorant about the detailed means of processing. Put differently, a deterministic system is fully transparent and traceable; a non-transparent system is not because of the opacity of the steps required to reach a specified output. Transparency in a technical system allows for easier tracking and reconstruction, which gains relevance in questions of responsibility (Mittelstadt et al. 2016) and influences social perception. Consequently, regarding certainty of output and traceability as crucial for human control consequently led us to the first and second dimensions of technical autonomy: non-transparency and indetermination.

As introduced above, the literature review revealed that adaptability is usually considered crucial for technical autonomy. We follow this position and consider adaptability to be the third core dimension. Being autonomous requires the capacity to learn and adapt behavior to a changing environment (Alonso and Mondragón 2004; Floridi and Sanders 2004; Pagallo 2017; Santosuosso and Bottalico 2017; Sartor and Omicini 2016; Thürmel 2015). A machine of this kind is able to process information, expand the knowledge implemented by programmers, and change the way it responds (Müller-Hengstenberg and Kirn 2016; Sartor and Omicini 2016; Thürmel 2015). This allows the system to adapt and to improve its performance in a certain environment (Sartor and Omicini 2016) without human intervention (Floridi and Sanders 2004; Pagallo 2017). Thus, adaptable systems are capable of altering their behavior which tends to make them more unpredictable for and independent from

human operators. Adaptability, therefore, shapes technical autonomy.

Interactivity is also regarded as a crucial feature of autonomy, as was revealed in the above discussion of core parameters. Advanced forms of interactivity result in system openness, what we consider the fourth key dimension when determining the level of autonomy. Various authors have recognized expansion of the original data through collaboration with other agents as a characteristic of autonomy and called this ability cooperation (Müller-Hengstenberg and Kirn 2016) or interactivity (Floridi and Sanders 2004; Pagallo 2017; Sartor and Omicini 2016). Due to collaboration, autonomous systems can delegate and divide labor among on another or build multi-agent systems (Müller-Hengstenberg and Kirn 2016). The shared ability to solve problems in a multi-agent system can exceed the capability of a single agent (Thürmel 2015; Weyer and Reineke 2005). The "intelligence" of multi-agent systems emerges through coordination among agents (Rammert and Schulz-Schaeffer 2002; Weyer and Reineke 2005) and hence is not predefined by programmers (Müller-Hengstenberg and Kirn 2016). It is not foreseeable with whom these technical systems will interact (Alonso and Mondragón 2004) and where they will gather their data, e.g., because the technical system is connected to other software agents. Openness with respect to such cooperation results in greater technical autonomy as does the openness and flexibility of the system with regards to data-gathering within the environment (and in other ways such as through the internet). Technical systems that are open in the sense that input is neither predefined nor limited are increasingly unforeseeable, resulting in the social ascription of more autonomy.

To sum up, due to a technical system's autonomy it is unforeseeable with whom it will interact and from where it will gather its data (openness). Furthermore, it can alter its behavior on the grounds of experience (adaptability). Its programming can leave it undetermined (indetermination) and untraceable (non-transparency). Our conceptualization thus identifies four dimensions defining the level of technical autonomy: non-transparency, indetermination, adaptability, and openness. Hence, to determine how autonomous a technical component is, one must answer the questions listed in Table 2.

This binary distinction is a simplification, and all of these characteristics can be present to varying degrees. However, this simplification is necessary to reduce complexity and make the taxonomy practical. Once the crucial features of the technical component are defined and the presence and absence of each dimension evaluated, the level of autonomy can be determined. The definitions of the taxonomy's levels of autonomy rely on the particular combination of features, as presented below and illustrated in Table 3.

4.3.1 Level 1

In the first and lowest level of autonomy, the technical component is transparent and determined, meaning that a given input always leads to a specified output, with full transparency with regards to how the system reaches that output. In such a machine, everything is completely predefined; the system is closed and has no ability to learn. The system is fully traceable and predictable. Simple calculators for example would operate on this level.

4.3.2 Level 2

A non-transparent technical system operates at the second level of autonomy. This means that the system remains fully

Table 2 Dimensions of technical autonomy with the respective assessment questions

Dimension	Assessment question
Non-transparency	Is the user and/or observer able to trace how the system gets from input A to output B?
Indetermination	Does input A always lead to the same output B?
Adaptability	Is the system able to learn from its experience?
Openness	Is the system able to expand its original input in particular due to cooperation with other systems and/or its flexibility in gathering source data?

Table 3 Levels of autonomy with an explanation of the main features

Level	Name	Description
1	Deterministic system	Technical system is determined, transparent, unadaptable, and closed
2	Non-transparent system	Technical system is determined, non-transparent, unadaptable, and closed
3	Indetermined system	Technical system is indetermined, non-transparent, unadaptable, and closed
4	Adaptable system	Technical system is indetermined, non-transparent, adaptable, and closed
5	Open system	Technical system is indetermined, non-transparent, adaptable, and open

determined but not every step is predefined and traceable. This marks an increase in autonomy, because the component potentially alters its manner of moving from the input to the output, and thus is non-transparent in its procedures. It holds back information and becomes opaque to the human operator or observer, which impacts social perception. However, such a system's output is still determined, because a certain input always leads to a particular output. An example for this level is a system that weights different parameters when reaching a decision. While the same input variables always lead to the same result, the human operator cannot fully trace how the relevant parameters are weighted in each single case.

4.3.3 Level 3

At the third level of autonomy, the technical component is indetermined and non-transparent. How a given input leads to a particular output is untraceable. The component is even more opaque, because it does not provide complete information about the execution steps and alters the output. At this level, there is no certainty of the result (i.e., the output is not determined). The same input does not necessarily result in a particular output; the system is not predictable. A common Level 3 application is a chance factor that is purposefully implemented to vary the output. Likewise, programming that links the system to a specific (yet at the moment of programming unknown) external input variable is a common characteristic of such applications.

4.3.4 Level 4

At level four, the technical system is able to learn from its own data. Once a system can learn, it can no longer be completely determined or tracked, because the input and pathway for reaching the output may change. Due to machine learning, an intelligent system possesses a considerable amount of autonomy. Such a system is not only undetermined due to a specifically implemented and randomized variable, it is capable of changing its behavior base and becomes adaptable to the environment. The output and single steps toward execution are thus variable and can be permanently adapted. This again marks an increase in opacity as it hampers the comprehensibility of the system's action. The human part can no longer comprehend the criteria and circumstances by which the actions of the technical system are coordinated. A technical system drawing upon a machine learning algorithm would be an example for this level of autonomy.

4.3.5 Level 5

A system is classified as Level 5 if it is undetermined and non-transparent. The former implies that a given input must not lead to the same output, while the latter indicates that the way it reaches the output is not easily accessible by the observer. Furthermore, it is adaptable due to its ability to learn, leaving room for modifications in the way it completes a task. It is capable of interacting with other systems and able to collect data. This openness marks the highest level of autonomy, because the system is neither limited to a specified and predefined input, nor to its own experience. Data gathering for the system's input is not under the full control of a human operator or programmer. It is, therefore, even more difficult to trace a system's actions, because those actions are opaque. A technical system based on a machine learning algorithm and additionally connected to an internet database, where it has access to new learning data would be an example for this fifth level of technical autonomy.

5 Discussion

The goal of this article was to introduce a tool that would allow everyone, regardless of technical expertise, to capture the basic structure of any human–machine collaboration. Two parameters lie at the core of our taxonomy that essentially characterize human–machine collaboration: automation and autonomy. We propose that every human–machine collaboration is typified mainly and most fundamentally by the nature of this cooperation (i.e., the distribution of tasks between the human and technical component) and the technical components' abilities and capacities, in other words, by the levels of automation and technical autonomy. Figure 1 maps these two axes and depicts the proposed taxonomy.

This taxonomy unites two research streams that run parallel to one another: the human factor and automation research (e.g., Endsley 1987; Parasuraman et al. 2000; Vagia et al. 2016), and technical sociology (Rammert and Schulz-Schaeffer 2002; Thürmel 2015) and computer ethics (Floridi and Sanders 2004) and thus rests on the concept of technical autonomy. Approaches that focus either on automation or on technical autonomy are not sufficiently differentiated. Only the combination of the two dimensions allows for a thorough understanding of human–machine collaboration, as both pose fundamental challenges. Understanding the level of automation is important, because the consequences are potentially far reaching; the allocation of control, among other aspects, affects the distribution of agency and, thus, responsibility (Chinen 2016; Kirchkamp and Strobel 2019). The higher the level of automation, the less human control there is over the system's actions. Thus, increasing levels of automation ask for higher requirements for example regarding precaution measurers, otherwise potentially resulting in liabilities. Conversely, discerning the level of autonomy is crucial, because this affects traceability and comprehensibility (Zarsky 2016). The greater the level of autonomy, the opaquer and more intractable the machine's actions appear.

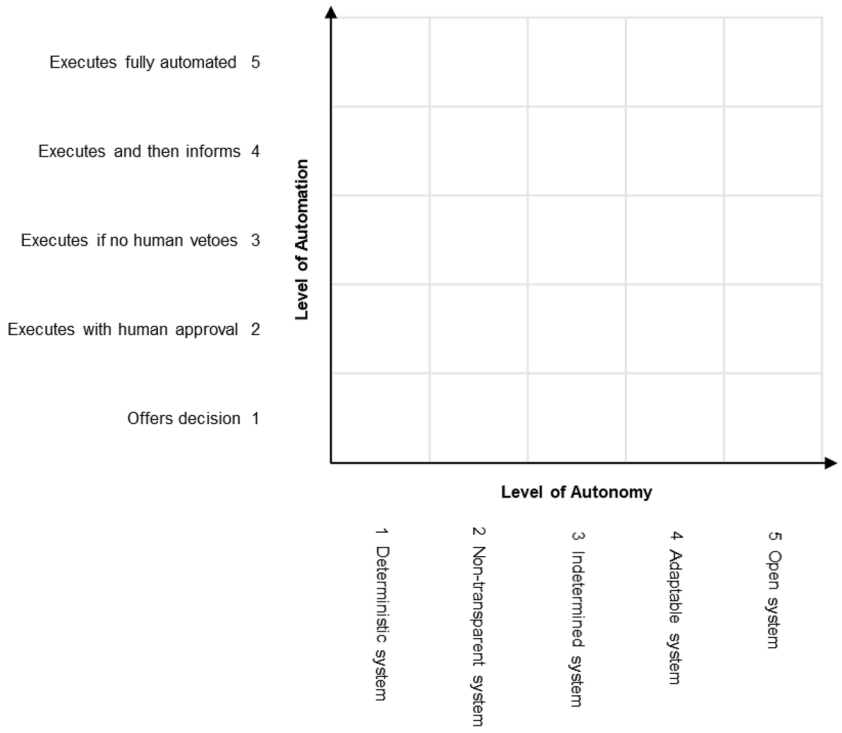

Fig. 1 Taxonomy of automation and technical autonomy in human–machine collaboration

This necessitates greater requirements in terms of comprehensibility and monitoring, as the human part of the system should still be able to understand and explain what the technical part has done. Challenges regarding a lack of explainability arise, likewise affecting questions of legitimacy (Mittelstadt et al. 2016), governance (Danaher et al. 2017) and accountability (Shin and Park 2019).

What a taxonomy does or does not capture is a chosen and, therefore, variable specification. Depending on the perspective, this specification may change. Automation and autonomy are not the only possible dimensions for describing advanced technology. There are, for example, taxonomies that classify socio-technical systems according to their degree of automation and complexity (Janssen and Kuk 2016). However, complexity is a consequence of advanced technology, and, therefore, not necessarily a core element of human–machine collaboration. Other scholars have concerned themselves with (technical) agency per se (Rammert 2009; Thürmel 2015). Although their analyses contribute very important and fundamental aspects that significantly influence this work, these studies have not allowed for a schematic assessment of the division of agency in terms of the concrete components determining the distribution itself. Other conceptualizations match different types of autonomy with different types of control in human-technology interaction, concluding that only humans exert the highest levels (Gransche et al. 2014). This assignment of different forms of autonomy and control to humans and technology is useful. It does not, however, explain in detail what characteristics and capabilities shape technical autonomy. Furthermore, as has been referred to throughout this work, there are many taxonomies and schemes that center solely on automation (see, e.g., Vagia et al. 2016). These neglect a dimension crucial to social perception, and, therefore, to many implementations.

Both, the dimensions and number of levels, depend on the perspective and particular taxonomy's aims. For example, scholars have reached different conclusions about the number of levels of automation (Vagia et al. 2016) or of autonomy (Gransche et al. 2014) necessary for adequate differentiation. Defining the appropriate number of levels is not, however, the most fundamental concern, as that number reflects only precision and usability. What matters more is that automation and autonomy are understood as gradual. While this has long been the case for automation (e.g., Endsley 1987; Sheridan and Verplank 1978), the understanding of technical autonomy as gradual is still relatively new (cf. for agency see Rammert 2009). Understanding the gradually of both dimensions contributes to a more differentiated capturing of advanced technology. Furthermore, it helps to more precisely highlight differences among human–machine collaborations.

5.1 Limitations

This taxonomy only allows for the classification of socio-technical collaboration within the boundaries of functional fulfillment. When evaluating such systems, the levels of automation and technical components' autonomy must be

specified. However, it is equally important to determine the function that a particular system fulfils (i.e., what task it is assigned). Appraising the legitimacy of the technical system requires careful consideration of whether the task is, in the first place, suitable for automation and autonomization. However, this feature cannot be captured abstractly. Therefore, this consideration cannot be part of the general taxonomy, but rather serve to complement an assessment.

This taxonomy also captures technical autonomy only in terms of what is possible today. It might well be that further technological developments lead to new and more advanced technological systems. However, if this is the case, our taxonomy can easily be extended and refined. It is important to note that this study's philosophical touchstones are not a limitation. In developing this taxonomy, concepts with substantial philosophical loading were unavoidable. Autonomy, as well as the states of being determined (and determinism) and undetermined (and indeterminism) are often used in technical literature. Employing them here should not lead to the misunderstanding that technology can assume these demanding philosophical requirements in the same ways as humans.

5.2 Avenues for further research

This taxonomy offers a number of initial points for future research. From a legal science perspective, investigating the influence of different levels of automation and autonomy on the attribution of legal responsibility is an interesting endeavor. Likewise, empirically testing if and to what effect the different levels have on punishment would be a fruitful field of research. From a management perspective, future work should derive implementation standards and precaution measures from this taxonomy, as well as describe requirements for organizations to use when implementing socio-technical collaboration. One example is a division into three levels, distinguishing between low-, medium-, and high-level implementation of human–machine collaboration. This categorization follows the assumption that a low level of autonomy or automation would lead to a low-level implementation, which in turn would require less monitoring and fewer precautions. Conversely, high values on both axes would indicate a high-level implementation, which would potentially be more demanding. Future research could also draw attention away from the collective and towards the individual level. Scholars could explore whether these levels have an actual effect on social and psychological perception, and if so, how this perception changes if automation and autonomy increase. One could also investigate what influence these levels have on decision-making and how they affect trust. In sum, there is a wide range of psychological, sociological, legal, and managerial research questions that could be explored based on this taxonomy, which will serve as a tool for clearly differentiating among the varieties of distributed agency in human–machine collaboration. Moreover, these findings will have a direct impact on policy discussions accompanying the implementation of technology in the digital age.

6 Conclusion

As was stated at the beginning of this work, the spread of advanced technology has resulted in human–machine collaboration becoming more and more prevalent. The consequences of their spread are far-reaching. Describing and understanding human–machine interaction is, therefore, crucial and should no longer be a task for engineers, software developers, or systems designers; it is now important for professionals and researchers in a wider variety of fields. However, socio-technical constellations are complex, and capturing them adequately calls for a distillation to their basic characteristics involved in human–machine collaboration: automation and autonomy.

Introducing this taxonomy based on these two dimensions may at first glance seem bold. It is important to note, though, that this boldness adds value. It allows the user to grasp complex phenomena and focus on what is most important in human–machine collaboration: the question of who does what on one hand, and the question of how independent is it done on the other hand. This clear description of the distribution of agency will allow professionals and researchers from different fields to estimate and evaluate the implications and consequences of given socio-technical constellations. We do not seek to oversimplify the discussion of socio-technical collaboration. To the contrary, by introducing different levels of automation and autonomy, we emphasize that these concepts are not a question of all or none, but rather vary in degree. The gradual approach presented here allows for a more differentiated description and understanding. Thus, the taxonomy points out that human–machine collaboration is multi-faceted, while also allowing for all users, despite their level of specialized knowledge, a pragmatic means of capturing them.

References

Alonso E, Mondragón E (2004) Agency, learning and animal-based reinforcement learning. In: Nickles M, Rovatsos M, Weiss G (eds) Agents and computational autonomy – potential risks and solutions. Springer, Berlin, pp 1–6

Balkin JM (2015) The path of robotics law. 6 California Law Review, Circuit 45.

Beck S (2015) Technisierung des Mensche: Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept "Verantwortung". In: Gruber MC, Bung J, Ziemann S (eds)

Autonome Automaten: Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft. BWV Verlag, Berlin, pp 173–187

Beer JM, Fisk AD, Rogers WA (2014) Toward a framework for levels of robot autonomy in human-robot interaction. J Hum Robot Interact 3:74–99

Bradshaw JM, Feltovich PJ, Jung H, Kulkarni S, Taysom W, Uszok A (2004) Dimensions of adjustable autonomy and mixed-initiative interaction. In: Nickles M, Rovatos M, Weiss G (eds) Agents and computational autonomy: potential, risks, and solutions. Springer, Berlin, pp 17–39

Castelfranchi C, Falcone R (2004) Founding autonomy: The dialectics between (social) environment and agent's architecture and powers. In: Nickles M, Rovatos M, Weiss G (eds) Agents and computational autonomy: potential, risks, and solutions. Springer, Berlin, pp 40–54

Chinen MA (2016) The co-evolution of autonomous machines and legal responsibility. Va J Law Technol 20:338

Danaher J, Hogan MJ, Noone C, Kennedy R, Behan A, De Paor A et al (2017) Algorithmic governance: developing a research agenda through the power of collective intelligence. Big Data Soc 4:1–21. https://doi.org/10.1177/2053951717726554

Endsley MR (1987) The application of human factors to the development of expert systems for advanced cockpits. Proc Hum Factors Soc Annu Meet 31(12):1388–1392. https://doi.org/10.1177/154193128703101219

Flemisch F, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J (2012) Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. Cogn Technol Work 14:3–18. https://doi.org/10.1007/s10111-011-0191-6

Floridi L, Sanders JW (2004) On the morality of artificial agents. Mind Mach 14:349–379. https://doi.org/10.1023/b:mind.0000035461.63578.9d

Franklin S, Graesser A (1997) Is It an agent, or just a program?: a taxonomy for autonomous agents. In: Müller JP, Wooldridge MJ, Jennings NR (eds) Intelligent agents III agent theories, architectures, and languages. ATAL 1996. Lecture notes in computer science (lecture notes in artificial intelligence). Springer, Berlin, pp 21–35

Gransche B, Shala E, Hubig C, Alpsancar S, Harrach S (2014) Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen. Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. Fraunhofer Verlag, Stuttgart

Hertzberg J (2015) Technische Gestaltungsoptionen für autonom agierende Komponenten und Systeme. In: Hilgendorf E, Hötitzsch S (eds) Das Recht vor den Herausforderungen der modernen Technik. Nomos, Baden-Baden, pp 63–74

Hilgendorf E (2017) Automated driving and the law. In: Hilgendorf E, Seidel U (eds) Robotics, autonomics, and the law. Nomos, Baden-Baden, pp 171–194

Janssen M, Kuk G (2016) The challenges and limits of big data algorithms in technocratic governance. Gov Inf Q 33:371–377. https://doi.org/10.1016/j.giq.2016.08.011

Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349:255–260. https://doi.org/10.1126/science.aaa8415

Kaber DB (2018) Issues in human–automation interaction modeling: presumptive aspects of frameworks of types and levels of automation. J Cogn Eng Decis Mak 12:7–24. https://doi.org/10.1177/1555343417737203

Kirchkamp O, Strobel C (2019) Sharing responsibility with a machine. J Behav Exp Econ 80:25–33. https://doi.org/10.1016/j.socec.2019.02.010

Korsgaard CM (2014) The normative constitution of agency. In: Vargas M, Yaffe G (eds) Rational and social agency: the philosophy of Michael Bratman. Oxford University Press, New York, pp 190–215

Lambe P (2007) Organising knowledge: taxonomies. Knowledge and organisational effectiveness. Chandos, Oxford

Loh W, Loh J (2017) Autonomy and responsibility in hybrid systems. In: Lin P, Jenkins R, Abney K (eds) Robot ethics 2.0: from autonomous cars to artificial intelligence. Oxford University Press, Oxford. https://doi.org/10.1093/oso/9780190652951.003.0003

Martin K (2018) Ethical implications and accountability of algorithms. J Bus Ethics 160:835–850. https://doi.org/10.1007/s10551-018-3921-3

Matthias A (2004) The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics Inf Technol 6:175–183. https://doi.org/10.1007/s10676-004-3422-1

Misselhorn C (2015) Collective agency and cooperation in natural and artificial systems. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-15515-9_1

Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. Big Data Soc 3:1–21. https://doi.org/10.1177/2053951716679679

Müller-Hengstenberg CD, Kirn S (2016) Rechtliche Risiken autonomer und vernetzter Systeme: eine Herausforderung. Walter de Gruyter GmbH, Berlin

NHTSA (2013) Preliminary statement of policy concerning automated vehicles. US National Highway Traffic Safety Administration, 30 May 2013

Nickerson R, Muntermann J, Varshney U, Isaac H (2009) Taxonomy development in information systems: developing a taxonomy of mobile applications. https://halshs.archives-ouvertes.fr/halshs-00375103/document. Accessed 3 Aug 2019

Nof SY (2009) Automation: what it means to us around the world. In: Nof S (ed) Springer handbook of automation. Springer, Berlin, pp 13–52

Nunes I, Jannach D (2017) A systematic review and taxonomy of explanations in decision support and recommender systems. User Model User Adapt Interact 27:393–444. https://doi.org/10.1007/s11257-017-9195-0

Onnasch L, Maier X, Jürgensohn T (2016) Mensch-Roboter-Interaktion-Eine Taxonomie für alle Anwendungsfälle. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Dortmund

Pagallo U (2017) From automation to autonomous systems: a legal phenomenology with problems of accountability. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17), pp 17–23. https://doi.org/10.24963/ijcai.2017/3

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern Part A Syst Hum 30:286–297. https://doi.org/10.1109/3468.844354

Proud RW, Hart JJ, Mrozinski RB (2003). Methods for determining the level of autonomy to design into a human spaceflight vehicle: a function specific approach. NASA Johnson Space Center Report, NASA Road, Houston, TX, 2003

Rammert W (2009) Hybride Handlungsträgerschaft: Ein soziotechnisches Modell verteilten Handelns. In: Herzog O, Schildhauer T (eds) Intelligente Objekte. Springer, Berlin, pp 23–33

Rammert W, Schulz-Schaeffer I (2002) Technik und Handeln: wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. In: Rammert W, Schulz-Schaeffer I (eds) Können Maschinen handeln?: soziologische Beiträge zum Verhältnis von Mensch und Technik. Campus Verlag, Frankfurt, pp 11–64

Rijsdijk SA, Hultink EJ, Diamantopoulos A (2007) Product intelligence: its conceptualization, measurement and impact on consumer satisfaction. J Acad Mark Sci 35:340–356. https://doi.org/10.1007/s11747-007-0040-6

Riley V (1989) A general model of mixed-initiative human-machine systems. Proc Hum Factors Soc Ann Meet 33:124–128. https://doi.org/10.1177/154193128903300227

Russell SJ, Norvig P (2014) Artificial intelligence: a modern approach. Pearson education limited, Malaysia

Santosuosso A, Bottalico B (2017) Autonomous systems and the law: why intelligence matters. In: Hilgendorf E, Seidel U (eds) Robotics, autonomics, and the law. Nomos, Baden-Baden, pp 27–58

Sartor G, Omicini A (2016) The autonomy of technological systems and responsibilities for their use. In: Bhuta N, Beck S, Geiss R, Lui HY, Kress C (eds) Autonomous weapon systems: law, ethics, policy. Cambridge University Press, Cambridge, pp 39–74

Sheridan TB, Verplank WL (1978). Human and computer control of undersea teleoperators. Institute of Technology Cambridge, Cambridge. https://www.dtic.mil/dtic/tr/fulltext/u2/a057655.pdf. Accessed 23 May 2019

Shin D, Park YJ (2019) Role of fairness, accountability, and transparency in algorithmic affordance. Comput Hum Behav 98:277–284. https://doi.org/10.1016/j.chb.2019.04.019

Shneiderman B (2016) The dangers of faulty, biased, or malicious algorithms requires independent oversight. Proc Natl Acad Sci USA 113:13538–13540. https://doi.org/10.1073/pnas.1618211113

Simmler M (2019) Maschinenethik und strafrechtliche Verantwortlichkeit. In: Bendel O (ed) Handbuch Maschinenethik. Springer, Wiesbaden, pp 1–18

Sommerville I (2007) Software engineering. Pearson Education Limited, Essex

Thürmel S (2015) The participatory turn: a multidimensional gradual agency concept for human and non-human actors. In: Misselhorn C (ed) Collective agency and cooperation in natural and artificial systems. Springer, Cham, pp 45–60

Vagia M, Transeth AA, Fjerdingen SA (2016) A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? Appl Ergon 53:190–202. https://doi.org/10.1016/j.apergo.2015.09.013

Verhagen H (2004) Autonomy and reasoning for natural and artificial agents. In: Nickles M, Rovatsos M, Weiss G (eds) Agents and computational autonomy. Lecture notes in computer science, vol 2969. Springer, Berlin, pp 83–94

Wein LE (1992) Responsibility of intelligent artifacts: toward an automation jurisprudence. Harvard J Law Technol 6:103–154. https://heinonline.org/HOL/P?h=hein.journals/hjlt6&i=109. Accessed 8 Aug 2019

Weyer J (2006) Die Kooperation menschlicher Akteure und nichtmenschlicher Agenten: Ansatzpunkte einer Soziologie hybrider Systeme. Working Paper, 16-2006. Wirtschafts- und Sozialwissenschaftliche Fakultät Universität Dortmund, Dortmund, pp 1–36. https://nbn-resolving.de/urn:nbn:de:0168-ssoar-120992. Accessed 10 June 2019

Weyer J, Reineke S (2005) Creating order in hybrid systems: reflections on the interaction of man and smart machines. Working Paper, 7-2005. Technische Universität Dortmund, Dortmund, pp 1–48. https://nbn-resolving.de/urn:nbn:de:0168-ssoar-109749. Accessed 10 June 2019

Zarsky T (2016) The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. Sci Technol Hum Values 41:118–132. https://doi.org/10.1177/0162243915605575