Bus Inf Syst Eng 65(4):391–401 (2023)
https://doi.org/10.1007/s12599-023-00789-9

RESEARCH PAPER

Augmented Intelligence, Augmented Responsibility?

Nick Lüthi · Christian Matt · Thomas Myrach · Iris Junglas

Received: 1 July 2021 / Accepted: 23 December 2022 / Published online: 15 February 2023

Abstract Intelligence Augmentation Systems (IAS) allow for more efficient and effective corporate processes by means of an explicit collaboration between artificial intelligence and human judgment. However, the higher degree of system autonomy, along with the enrichment of human capabilities, amplifies pre-existing issues of the distribution of moral responsibility: If an IAS has caused harm, firms who have operated the system might argue that they lack control over its actions, whereas firms who have developed the system might argue that they lack control over its actual use. Both parties rejecting responsibility and attributing it to the autonomous nature of the system leads to a variety of technologically induced responsibility gaps. Given the wide-ranging capabilities and applications of IAS, such responsibility gaps warrant a theoretical grounding in an ethical theory, also because the clear distribution of moral responsibility is an essential first step to govern explicit morality in a firm using structures such as accountability mechanisms. As part of this paper, first the necessary conditions for the distribution of responsibility for IAS are detailed. Second, the paper develops an ethical theory of Reason-Responsiveness for Intelligence Augmentation Systems (RRIAS) that allows for the distribution of responsibility at the organizational level between operators and providers. RRIAS provides important guidance for firms to understand who should be held responsible for developing suitable corporate practices for the development and usage of IAS.

Keywords Responsibility gaps · Intelligence augmentation systems · Reason responsiveness · Algorithmic responsibility

Accepted after four revisions by Roman Beck.

N. Lüthi (✉) · C. Matt · T. Myrach
Department Information Management, Institute of Information Systems, University of Bern, Engehaldenstrasse 8, 3012 Bern, Switzerland
e-mail: nicola.luethi@unibe.ch

C. Matt
e-mail: christian.matt@unibe.ch

T. Myrach
e-mail: thomas.myrach@unibe.ch

I. Junglas
Department of Supply Chain and Information Management, College of Charleston, 66 George Street, Charleston, SC 29424, USA
e-mail: junglasia@cofc.edu

1 Introduction

More powerful algorithms provide better decision-making support and enable more automation and abstraction of human decisions through Intelligence Augmentation Systems (IAS). IAS are a specific form of autonomous system (Krenzer et al. 2019) that complement human intelligence to enable more efficient and more autonomous decision-making (Zhou et al. 2021). More system autonomy is often considered a trigger for process improvements (Galliers et al. 2017). However, the allocation of responsibilities between system providers and operators is difficult when it comes to IAS (Rohner 2013), especially due to the innate autonomy of IAS. More system autonomy does not necessarily create new forms of responsibility challenges but rather enhances pre-existing ones, as it is the system autonomy that allows IAS operators and providers to shift blame to the IAS. For instance, if recruiting software that uses augmented intelligence creates biased recommendations for new hires, who should be blamed for the biased

hiring practices resulting from this – the firm that operates the recruiting software but whose employees have little control over the IAS's functioning? Or maybe the firm that has provided the recruiting software, and whose employees implemented the IAS's underlying rules but who lacks control over how the system is eventually used by the operating firm?

IAS and their innate autonomy undermine the control that providers and operators previously had over technologies (Beck et al. 2022), thus creating so-called technological responsibility gaps (Johnson 2015). These are states that emerge after employees have made use of technology and in which it is impossible to attribute responsibility directly to the employees that have caused the current state because they had insufficient control over the technology. Such responsibility gaps are often inevitable consequences, particularly in complex technological scenarios involving multiple human and non-human actors (Matthias 2004). They might particularly emerge if there is only a distant relation between provider and operator (Lüthi et al. 2021), or if operators have little influence on the IAS design. This is typically the case for standard software, which provides a large basis of predefined functionality that caters to mass-market requirements and that needs to be configured by operating firms to account for their individual requirements. In this case, responsibility gaps are insidious because they are not ethically alarming in every situation in which they emerge. If no one is harmed, the distribution of responsibility might seem less urgent. However, besides preparing the ground for immoral actions, it also leaves firms in a critical state because a clear distribution of responsibility is a precondition for developing suitable means of corporate governance structures, including accountability mechanisms. After all, firms need to know first who should be blamed before effective measures for establishing explicit morality can be taken.

Research realized early on that technological settings could pose responsibility challenges (Khalil 1993) and affect individual responsibility perceptions (Harrington 1996). This was followed by discussions of how responsibility gaps can arise due to the use of technologies (Matthias 2004), as well as perceptions of technology as amoral can become reasons for the emergence of responsibility gaps and how adequate design processes can prevent these (Johnson 2015; Martin 2019a; Santoni de Sio and Mecacci 2021). However, while providing important groundwork, previous research still lacks a coherent ethical theory that allows for a clear distribution of responsibility for IAS when a responsibility gap occurs. Likewise, previous research does not account for the recent advancements in autonomy around IAS and the essential shifts in responsibility that higher degrees of autonomy entail. We believe it is an unsustainable condition that both providers and operators simply blame IAS-technology to escape their responsibility, only because no reference frame for distributing the responsibility exists. We therefore develop an ethical theory of Reason-Responsiveness for Intelligence Augmentation Systems (RRIAS) to show how responsibility for actions initiated by IAS can be distributed amongst IAS providers and operators.

While the actions that lead to harm do not have to be deliberate, RRIAS demonstrates that both providers and operators can manoeuvre an IAS's potential for harm through their design and use of system features. Their influence via design and usage options by means of system features leads to a certain level of implemented reason-responsiveness, i.e., in form of the resulting IAS's ability to account for moral reasons provided by employees during design or use and to react appropriately to these reasons. The non-utilization of one's possibilities to impact reason-responsiveness, in turn, leads to provider- and operator-specific responsibility gaps, which constitute the basis for their resulting moral responsibility. By constructing a conceptual chain from individual IAS system design and use decisions towards responsibility distribution, RRIAS describes whether providers or operators have the main responsibility as well as what the extent of the responsibility is. Illustrated by four scenarios, we use RRIAS to show what a sensible and less sensible IAS design and use can look like, and what the consequences of these use and design decisions by IAS providers and operators are.

RRIAS is a normative ethical theory. As such it does not predict or explain behaviour but instead proposes what actors should ideally do and how actors can be held responsible for acting in certain manners. While previously prevalent mostly in philosophical disciplines, normative theories have recently gained more attention in IS research, given the changing nature of interactions between employees and machines, which is enabled by the rising autonomy and power of underlying AI-based systems (Stahl 2012). Stahl (2012) differentiates four levels of normative research in IS (Fig. 1). Our theory is located at the third level–it is an ethical theory that tells us what firms should and should not do. Theories at this level justify explicit morality, such as corporate guidelines, to establish accountability for the use of IAS.

Our theory development follows the three-stage process suggested by Rivard (2014). We first detail the need for new theory based on the shortcomings of the current literature and develop the theoretical foundation upon which we will build our theory. Second, we describe our theory in detail, including its boundary conditions, how IAS providers and operators affect responsibility gap sizes, and how these efforts lead into the RRIAS theory. Third, we discuss how RRIAS contributes to theory and practice.

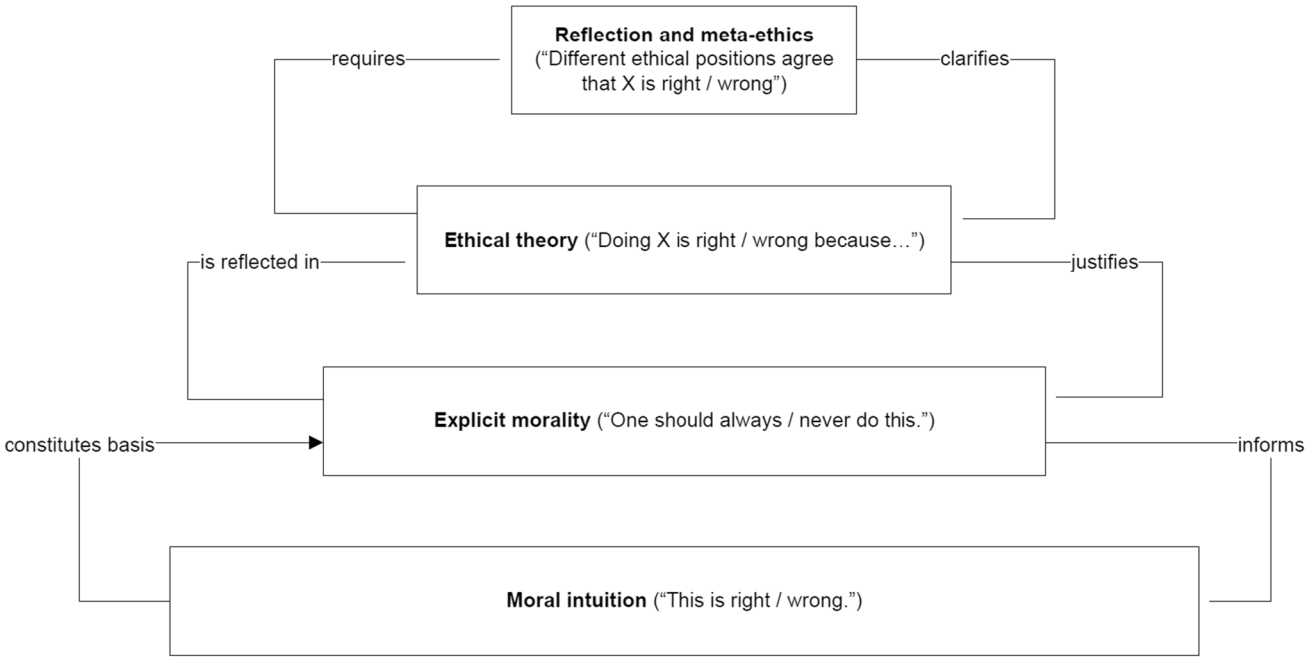

Fig. 1 The different levels of normativity (Stahl 2012, p. 638), (Copyright by B. C. Stahl)

2 Conceptual Background and Integration

As an ethical theory, RRIAS tells us what the ethically right thing is to do for firms when designing and using IAS to reduce their organizational responsibility facing potential harm. RRIAS distributes responsibility between two firms: the provider, and the operator of an IAS. Through designing and using the IAS and its system features, providers and operators influence the IAS's ability to react to moral reasons: its reason-responsiveness. By determining a system's reason-responsiveness, firms participate in influencing how large the responsibility gaps are that arise from its use; they are also morally responsible for the size of these gaps because they control an IAS's reason-responsiveness.

The following sections provide the foundation for our theory by detailing why IAS are specifically insidious and warrant an ethical theory, what exactly constitutes a firm's moral responsibility and how reason-responsiveness is controlled by IAS providers and operators. For this, we first outline a theoretical frame around causal and moral responsibility before we develop the concept of reason-responsiveness, which is a key pillar of RRIAS.

2.1 Intelligence Augmentation Systems

Following Zhou et al. (2021), we define IAS as partly autonomous systems that fulfil the specific function to augment human decision-making processes and capabilities. IAS have the function to "enhance[e] and elevat[e] human's ability, intelligence, and performance with the help of information technology. IA [Intelligence Augmentation] stresses human–machine collaboration or human–machine symbiosis where machines perform what they do best (e.g., computing, recording, and doing routine, repetitive work) to aid humans in doing what humans do best (e.g., abstract reasoning, creating, and making in-depth discoveries about people and the world)" (Zhou et al. 2021, p. 245). As such, IAS are technology agnostic; they are a subset of autonomous systems and can thus have varying levels of autonomy (Sheridan and Parasuraman 2005; Janiesch et al. 2019). IAS and other types of autonomous systems demand strategic choices of the businesses that employ them and, more importantly, such strategic choices are "highly consequential for a number of stakeholders" (Marabelli et al. 2021, p. 7). A growing part of the literature highlights ethical concerns that are directly applicable to IAS, such as the ethical issues of human–machine interaction (Stephanidis et al. 2019), specific algorithmic issues, such as bias (Kordzadeh and Ghasemaghaei 2021; Köchling et al. 2021), and organizational readiness for artificial intelligence (Jöhnk et al. 2021). We focus on situations in which human intelligence is augmented by an information system deployed as standard software, irrespective of the specific underlying technology.

2.2 Causal and Moral Responsibility

Responsibility is a broad, multi-faceted concept. The term originates from moral philosophy and is void of a

commonly accepted understanding (Pettit 2007; Johnson 2015; Fischer and Ravizza 1998). There is, however, widespread agreement to distinguish between two types of responsibility: *causal responsibility* and *moral responsibility* (Collins 2019). Causal responsibility simply denotes something's or someone's position within a causal chain of events (Sartorio 2016). For example, if a recruitment IAS used for hiring decisions makes a biased recommendation, then the IAS is causally responsible for the biased recommendation because it directly causes the recommendation. Yet, nobody will hold the recruitment IAS morally responsible because it lacks certain conditions for moral responsibility (Wallach and Allen 2008). Causal responsibility is considered a prerequisite for moral responsibility: If no one (or nothing) makes a biased recommendation, then no one needs to be held morally responsible for it. To hold someone causally responsible is a weaker claim than to hold someone morally responsible (Collins 2019). However, if, for example, a firm exhibits biased hiring practices, things change. Moral responsibility is an act of judgment that is ascribed to someone. If someone, in this case the firm, is blamed for actions, then it is because these actions were not the morally right thing to do (Collins 2019; Mason 2019). There is a relation between an object and a subject to which one ascribes moral responsibility (Stahl 2006). The firm is morally responsible because it should be blamed for its employees' actions that were done under the firm's name. To be morally responsible is to say that someone is more than just the culprit. Rather, this someone is a culprit who meets the threshold of blameworthiness for their actions (Levy 2005; Mason 2015).

There are numerous accounts of what exactly constitutes blameworthiness for moral responsibility. Our theory will follow Pettit (2007) which considers three conditions that collectives, such as firms, must fulfil to be considered blameworthy and thus to be held morally responsible:

- *Value relevance*: The firm is acting autonomously, facing "a value-relevant choice involving the possibility of doing something good or bad or right or wrong" (Pettit 2007, p. 175)
- *Value judgment*: The firm has the ability to make judgements by comparing the values of the different actions available to them.
- *Value sensitivity*: The firm has "the control necessary for being able to choose between options on the basis of judgments about their value." (Pettit 2007, p. 175)

The first two conditions are general in nature and relate to the acting firm. Specifically, the firm as a collective comprised of its employees, must be autonomous, have a choice between different possible possibilities, and have the mental capacity to distinguish between these possibilities. These two conditions manifest themselves in autonomous, fully conscious employees, and can thus also be presupposed in firms as a whole. The third condition concerns the situation in which an action is taken and is crucial to actions in the firm. For that, a firm collectively needs to have sufficient control over its actions to be responsible. The latter condition is therefore crucial for deciding whether and how IAS providers and operators are to blame. To be held morally responsible, a firm's employees need to have the necessary autonomy and control over actions done in the name of the firm in relation to an IAS (Kellogg et al. 2019). When it comes to using and designing information systems, there are established strategies that allow IAS providers and IAS operators to collectively divert and reject responsibility ascriptions (French 1984). For example, system providers argue that it is the IAS operator's responsibility to use the IAS correctly after purchase, while system operators argue that they were not given enough control over an IAS to avoid potential harm (Sparrow 2007; Hellström 2013).

2.3 Reason-Responsiveness

Employee autonomy as well as system autonomy are important factors for assigning responsibility and are particularly relevant for IAS (Newell and Marabelli 2015; Faraj et al. 2018). The innate autonomy of IAS diminishes the control firms and their employees have over the system's actions. To solve this potential control problem, Santoni de Sio and Hoven (2018) argue that IAS and other systems should track the relevant moral reasons of employees using the system as this would enable human control and oversight. Mecacci and Santoni de Sio (2020) extend this view by arguing that an IAS does not need a morality tracker to capture all reasons of employees; instead, the IAS should be reason-responsive, i.e., it should have the ability to account for *proximal* and *distal* moral reasons provided by employees during design and use, and to react appropriately to these reasons during use.

The reasons being provided to the IAS can be of varying distances. Mecacci and Santoni de Sio (2020) differentiate between distal and proximal reasons. *Distal reasons* are moral reasons that operate at a larger, societal scale, such as values, norms and plans, and are usually implemented during the design phase (Mecacci and Santoni de Sio 2020). For instance, an IAS that supports recruitment processes might incorporate the value "neutrality" and might force employees of the operator to provide neutral, non-biased data to continue. Such a reason is considered distant because it is inscribed into the IAS by the system provider without considering the specific corporate values of the IAS operator. Even if the operator holds different or differently weighted values as part of its corporate guidelines, its employees will still need to obey the enforcement

mechanisms as designed by the provider when using the system. In contrast, *proximal* reasons are closer to the specific use cases and are usually provided directly during use. Proximal reasons operate at a smaller scale and encompass the intentions behind the use case and the employees associated with the execution of a task (Mecacci and Santoni de Sio 2020). For example, if HR has the intention to conduct a fully neutral recruitment process but cannot exclude that the IAS suffers from certain biases, HR employees might use the proximal reason of "avoiding harm to anyone" by circumventing the system's recommendation and applying their own expertise and knowledge in a manual ranking process instead. Such reasons are proximal because they can vary between employees and use cases. This is an important distinction from distal reasons, which are agreed upon on at a collective level; they can be firm-wide (i.e., each IAS provider has specific distal reasons for their products), society-wide, or even global.

An IAS following Mecacci and Santoni de Sio's (2020) account would not necessarily track all relevant moral reasons, but it would respond to relevant moral reasons if warranted. In other words: It is up to the IAS providers and operators to determine the relevant moral reasons – proximal or distal – and how to handle them. For a recruitment IAS, the provider might decide that "neutrality" is a core value to which the IAS should be responsive to (i.e., distal reason-responsiveness), while the operator decides whether and how their employees can provide their own proximal reasons (i.e., proximal reason-responsiveness), for example, by allowing employees to inspect all applicants if they find the recommendations unfair.

To create a system following these requirements, Mecacci and Santoni de Sio (2020, p. 112) propose the following design approach: "(i) respond to a proximal reason IFF [if and only if] it does not conflict with a more distal reason, and (ii) respond to the most proximal reason allowed by (i)." This approach can also be perceived as a first preliminary ethical theory that distributes responsibility in an algorithmic setting. Specifically, and as part of such an ethical theory, blame can be assigned in the following fashion: IAS providers can be blamed for choosing inappropriate distal reasons to be inscribed into the IAS, or for not allowing relevant proximal reasons to be aptly considered in cases of conflict. IAS operators can be blamed for selecting an IAS that makes use of inappropriate distal reasons, or distal reasons not aligned with theirs, or if they do not execute on the proper proximal reasons.

3 Developing a Theory of Reason-Responsiveness for Intelligence Augmentation Systems

Key actors of our Theory of Reason-Responsiveness for Intelligence Augmentation Systems (RRIAS) are providers and operators of Intelligence Augmentation Systems who control the reason-responsiveness of an IAS. To hold firms morally responsible, RRIAS needs to show how firms that develop/use IAS as standard software fulfil the third condition of moral responsibility, i.e., how it exercises control over the IAS. To achieve this, the following section will provide the necessary arguments in two steps.

We first provide the model of our ethical theory and argue how IAS providers and operators by implementing reason-responsiveness through the use and design of system features control the size of the resulting responsibility gap, rendering them responsible. Second, we will provide four scenarios that illustrate how the amount of control varies between IAS providers and IAS operators and how RRIAS can be used to distribute responsibility in these scenarios. The Online-Appendix summarizes the definitions of the core elements of RRIAS and provides an example as an illustration for each.

3.1 Conceptual Model

Our conceptual model details the causal chain of actions, from actors to a system's reason-responsiveness, and summarises how the RRIAS theory distributes moral responsibility for providers and operators (Fig. 2).

The model includes IAS providers and operators as actors. The provider develops and sells the IAS as standard software; the operator purchases, configures, and uses it. Each firm comprises a collective of employees fulfilling the conditions for moral responsibility. As a firm, they each have a collective responsibility for the development/sales of the IAS (provider) and a collective responsibility for the purchasing/use of the IAS (operator). The developed/used IAS is at least partially autonomous, but employees always make the final decision (i.e., either accept or reject an IAS's suggestions).

An IAS provider designs the system features of an IAS which determine its distal reason-responsiveness and enable its proximal reason-responsiveness. An IAS operator then uses the system features and determines its proximal reason-responsiveness through use. Reason-responsiveness (distal and proximal) ranges from high to low and directly impacts the size of the resulting responsibility gap. By designing and using system features, providers and operators execute control over the reason-responsiveness of the IAS they use. The more actions are executed without proper control, that is actions that are not governed by either intentional proximal or distal reason-

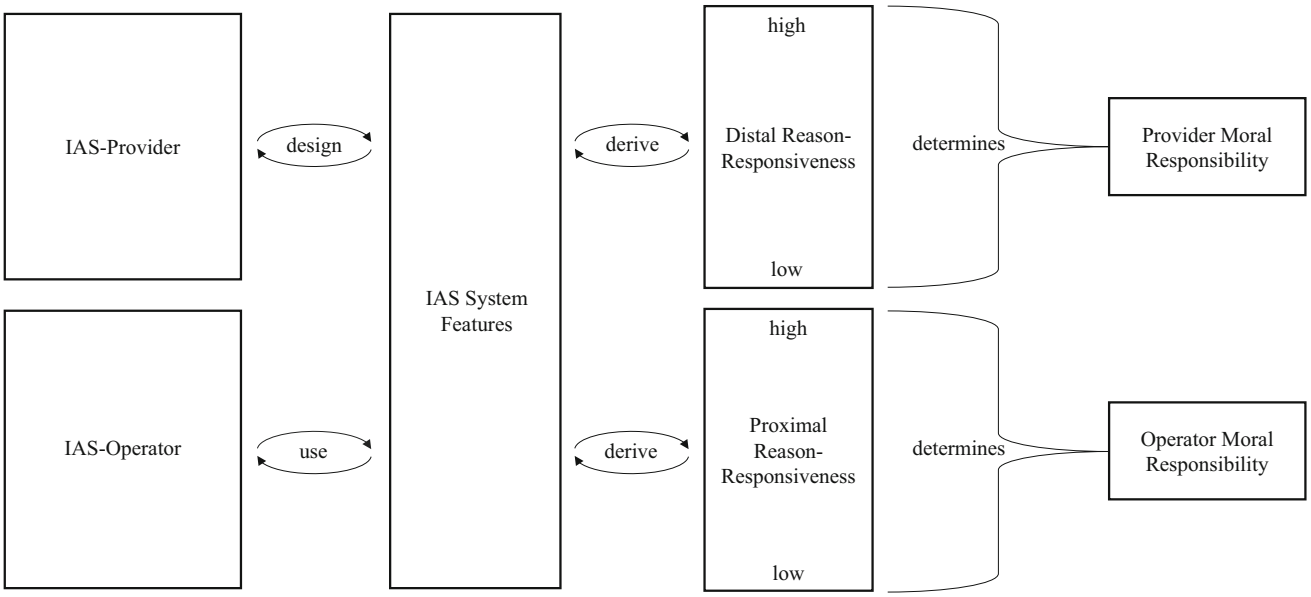

Fig. 2 Conceptual RRIAS model

responsiveness, the larger the resulting responsibility gap. The size of the responsibility gap is relevant because it corresponds to the provider's and operator's responsibility. As our discussion of moral responsibility has shown, the deciding factor for moral responsibility (employee autonomy and agency assumed) is the amount of exercised control. By controlling a system's reason-responsiveness, providers and operators can be blamed for failing to do so properly.

It is important to note that both the design and use of the system features are processes. The system's features and its reason-responsiveness influence each other. Some system features might be designed and used without the provider's intent to influence its reason-responsiveness but might still heavily affect it. Conversely, considerations about a system's reason-responsiveness might affect system features and how they are implemented into the system. System features and reason-responsiveness will change over time.

3.2 Distributing Responsibility with RRIAS

Using four scenarios, we detail how RRIAS enables the distribution of responsibility between system provider and operator. The scenarios are purposefully selected to illustrate and discuss cases of explicit misuse and inadequate design at a corporate level and by outlining IAS operators and providers' influence in affecting distal and proximal reason-responsiveness through design and usage decisions at the feature level. Such misuse and inadequate design can arise for many reasons and are not necessarily attributed to bad faith. Providers might inadequately design an IAS because markets demand quick development times that require them to reduce testing cycles, or because the firm's guidelines and processes are not equipped for ethical considerations. Operators might misuse IAS because they might simply adopt usage guidelines and practices of a previous system without accounting for the complexity of the IAS, or because they fail to instruct employees on the specificities of the IAS.

All four scenarios, detailed in the following, involve an IAS operator that has recently purchased a recruitment IAS (developed by the IAS provider) as standard software and that applies augmented intelligence to screen job applicants. The IAS collects all applications, assigns each applicant a score, and filters the top 10% of applicants sorted by score. The IAS is intended to be used only for screening job applicants, not for the internal scoring of employees or any other uses. Table 1 provides an overview of the four scenarios, the adequacy of design and use, the resulting reason-responsiveness, and responsibility gaps.

For the well-intentioned scenario, we assume that both provider and operator have conducted their due diligence and operate under the best of intentions. As such, the provider will try to make the system highly responsive to distal reasons and enable high reason-responsiveness through the system's features and configuration options given to the operator. The operator will ensure high proximal reason-responsiveness when using the system. The hiring system thus would look as follows:

> *Well-intentioned:* The IAS provider designs the IAS with two additional features: If the best scored 10% of the applicants are too similar (e.g., with cumulation of one gender or educational background), the IAS issues a warning and shows a decision tree detailing how

Table 1 Four scenarios detailing the distribution of responsibility according to RRIAS

Scenario	Design/use	Reason-responsiveness	Responsibility gap
Well-intentioned	Well-intentioned design	Distal: high	Small for both
	Well-intentioned use	Proximal: high	
Misuse	Well-intentioned design	Distal: high	Larger for operator
	Misuse	Proximal: low	
Inadequate design	Inadequate design	Distal: low	Larger for provider
	Well-intentioned use	Proximal: high	
Complete negligence	Inadequate design	Distal: low	Large for both
	Misuse	Proximal: low	

parameters were weighed (warning feature). Furthermore, the provider provides three mechanisms if a warning is shown between which the operator can choose during implementation (choice feature): 1. The warning is ignored and not displayed to employees. The system outputs the score. 2. The scores are kept but the threshold of applicants shown is lowered (from top 10% to top 20%). 3. The scores are hidden from the employees and all applicants are shown, and sorted alphabetically. The operator decides to implement the third mechanism, showing all applicants to employees.

In this scenario, the warning feature incorporates distal reasons, namely that members of society should be treated equally without any regard for parameters, such as gender or race, and the system can be rendered transparent if the need arises. However, by providing different implementation choices to the operator, the provider does not enforce these reasons, but rather enables proximal reason-giving through implementation and use. The operator decides to which degree it wishes to implement proximal reason-responsiveness for its employees: not at all (ignore warnings), to a certain extent (lower threshold) or fully (no scores), subsequently deciding on a full implementation of proximal reason-responsiveness. The resulting responsibility gaps for both proximal and distal reason-responsiveness are small, the provider has incorporated distal reason-responsiveness as well as enabled proximal reason-responsiveness, while the operator ensures proximal reason-responsiveness by enabling its employees to see all applicants and use their expertise, or proximal reasons, to decide on the outcome. Consequently, each firm's collective responsibility is considerably reduced as the system is designed and used with well-intentioned features. Ideally, every IAS is built and used according to this well-intentioned scenario: Providers implement distal reasons and enable operators to provide proximal reasons, and the operator enforces the use of proximal reasons amongst its employees when using the system. The resulting responsibility gap, assuming a continued well-intentioned design and use, is small for both firms.

For the misuse scenario, we assume that the provider has done its due diligence and built a system accordingly by enabling distal and proximal reason-responsiveness. The operator, however, uses the system for a purpose it was not intended for:

Misuse: The IAS provider designs the IAS again with both, the warning feature as well as the choice feature. The operator makes use of the choice feature and decides to ignore the warning feature and to always provide a scoring to the employees, irrespective of whether the system issues a warning or not, as this streamlined approach allows for more standardized processes. Additionally, after using the system for screening job applicants and due to its success in that task, the operator decides to also use the system's score as a mandatory metric for internal career advancement decisions.

The provider-induced responsibility gap in this scenario is small because it has built a well-intentioned system and thus has done what is in its power to reduce responsibility by making the system responsive to distal reasons (equality, transparency) and by enabling proximal reason-responsiveness. The operator on the other hand has done little to reduce the resulting responsibility gap. By not implementing the warning feature, it does not give its employees the possibility to provide their proximal reasons during use. More gravely, by misusing the system, the operator circumvents the distal and proximal reason-responsiveness provided by the system's features. We cannot (and should not) reasonably expect IAS providers to make their systems reason-responsive to any possible scenario, but rather reason-responsive for scenarios the system is intended for. By misusing the system (i.e., not using it for its intended goal), the operator is effectively using a system that is not reason-responsive at all. Responsibility thus falls squarely on the operator.

For the inadequate design scenario, we assume that the provider has done little to reduce distal reason-responsiveness but enables proximal reason-responsiveness:

Inadequate Design: After receiving the relevant data of all applicants, the IAS issues a score for each entry. Because several similar systems were recently released, the provider is adamant to protect its proprietary scoring algorithm. After some internal discussion, the provider thus decides against incorporating the warning and choice feature amidst fears that those features might reveal crucial information about the inner workings of the algorithm. The operator is consequently not given any significant choices when implementing the algorithm. Its employees do not have the possibility to understand how the score was derived and what criteria were used. The operator thus decides to only use the recruitment system in a supporting function and for jobs with many applicants.

In this scenario, the IAS is not reason-responsive. No moral reasons (distal or proximal) were considered when designing the system and its features. However, the operator was aware of that and tried to implement the system in such a way that its lack of reason-responsiveness is offset by how it is used. The resulting responsibility gap for the provider is large, while it is small for the operator. Almost all the responsibility lies with the IAS provider. Its design choices for the system features did not incorporate distal reasons, nor did it enable proximal reason-responsiveness (or only by accident). If at all, the operator can only be blamed for using the system. By making adequate implementation choices, it has done as much as possible to retain a certain amount of control over the system.

For the complete negligence scenario, both provider and operator do nothing to enhance the system's proximal and distal reason-responsiveness:

Complete Negligence: After receiving the relevant data of all applicants, the IAS issues a score for each entry. Because several similar systems were recently released, the provider is adamant to protect its proprietary scoring algorithm. After some internal discussion, the provider thus decides against building the warning and choice feature amidst fears that those features might reveal crucial information about the inner workings of the algorithm. The operator, although aware of these limitations, decides to not make significant changes when implementing the system as it deems the system's algorithm to be quite good. As part of standardization efforts, it mandates its employees to rely on the choice of applicants provided by the system.

In this scenario, the IAS is not reason-responsive. No moral reasons (distal or proximal) were considered when designing and using the system and its features. The resulting responsibility gap for both provider and operator are thus considerably large. The IAS provider is responsible because its design choices for the system features did not incorporate distal reason-responsiveness, nor did it enable proximal reason-responsiveness (or only by accident). The responsibility gap is just as large on the IAS operator side. It is responsible for using a system that is not reason-responsive as well as for the implementation choices of the IAS, which might have allowed at least a certain degree of control.

4 Implications of RRIAS

4.1 Theoretical Contributions and Implications

IAS governance has received little attention despite the current surge of research on IAS in general (Zhou et al. 2021). RRIAS provides the underpinning for the ethical governance of IAS by allowing for the distribution of responsibility between providers and operators in settings where IAS are used as standard software. Our ethical theory shows that the strategies of blame-shifting between providers and operators are not ethically sound. RRIAS shows that both firms have clear responsibilities to develop and use IAS that account for distal and proximal reason-responsiveness. As we have shown, the distribution of responsibility is closely connected to the processes of designing and using IAS. There is no universal solution to the distribution of responsibility, therefore it should rather be distributed on a case-by-case basis. As such, the deliberations of RRIAS can serve as a basis for more complex cases, extending standard software use, such as multiple firms designing and using an IAS, the involvement of subsidiaries, etc. RRIAS also can serve as a basis for more autonomous systems, such as fully autonomous decision-making or other means of artificial intelligence systems.

We extend the literature on technological responsibility gaps (or voids) (Collins 2019; Johnson 2006; Matthias 2004; Martin 2019a; Braham and VanHees 2011) in numerous ways. We proposed a definition of technological responsibility gaps, which has been lacking until now, and we show how IAS operators and IAS providers influence these gaps of responsibility. Our theory also allows for a normative distribution of responsibilities between IAS providers and operators. We thus improve the tangibility of ethical research about responsibility for IAS providers and operators. Our theory avoids the potential pitfalls when theorizing about responsibility gaps as discussed by Santoni de Sio and Mecacci (2021): Our account links responsibility to accountability and shows how one can normatively theorize about a certain subset of responsibility gaps. As such, our account is not fatalistic (too narrow), deflationistic (only focused on accountability), or solutionistic (too focused on technology as a meaningful

solution). Furthermore, our theory directly influences its underlying theories, proposed by Santoni de Sio and Hoven (2018) and Mecacci and Santoni de Sio (2020). We have taken the concept of reason-responsiveness and employed it as a determinant for the size of a responsibility gap while arguing that IAS providers and operators control this aspect independently of the IAS's actual autonomy. We extend the concept of meaningful human control and propose a more tangible approach with regard to responsibility for the design and use of IAS in a standard software setting.

Due to the normative nature of RRIAS, it advises what actors ideally should do. In the ideal case, IAS providers and operators would always act in accordance with the well-intentioned scenario and would thus try to create IAS with a high to very high reason-responsiveness to minimize their responsibility. In the not-so-ideal case, namely if both actors display complete negligence, they are to fully blame for IAS development and use. In practice, it will probably be much more common that firms encounter scenarios in a grey area, such as our "inadequate design" or "misuse" scenarios, in which one firm fails to adhere to its responsibilities as proposed by RRIAS. If we assume that the provider or operator falls short of its responsibilities and has no intent to change that, the normatively loaded question remains as to what the other firm should do: How should providers react if their IAS is misused? And how should operators act if an IAS they are using, or wish to use, is inadequately designed?

If providers build IASs with high reason-responsiveness, they have done as much as they can directly in their power and under their control. The brunt of misuse responsibility lies with the operator. However, there are different forms of misuse, and it makes a significant difference whether a single employee circumvents certain IAS functionalities or whether the IAS was sold to an operator that now uses it for unintended purposes. In the first case, there really is not much an IAS provider can or would do in practice. Assuming the provider offers adequate documentation or schooling to end users, its responsibility is very small according to RRIAS. In the second case, the answer is not as clear-cut. If the operator was clearly told what the IAS does and for which areas it should be used, the provider is generally not responsible for misuse. If the provider, however, sells to malicious operators, it should be blamed for that. Knowingly selling to an operator that will use the IAS for purposes other than the ones intended decreases the reason-responsiveness of the IAS. Thus, providers should not only build their IASs as reason-responsive as possible but they should also clearly communicate to operators the boundaries of the IAS. The responsibilities of operators are slightly different. Generally, operators are responsible for what IAS they use and how they use it. Ideally, operators never make use of inadequately designed IAS. It is thus vital for operators to ensure that an IAS is built for their specific use case as well as that they vet any IAS they use for potential issues before they are used in production environments. There might be cases in which operators have no choice but to use a specific IAS even though they are aware of its issues, for example, due to regulatory requirements or legacy issues. In these cases, operators need to adapt their use case to the shortcomings of the IAS; their users should be informed about the IAS's issues and its use should be guided by clear boundaries and guidelines to minimize the potential of issues arising due to the IAS's inadequate design. As long as the use of an IAS is a free choice made by operators, they are always partially responsible for the use of an IAS.

4.2 Practical Implications

Our theory allows both IAS providers and operators to determine to which degree they are responsible. RRIAS enables actions at a firm-level to reduce responsibility gaps and to ameliorate the reason-responsiveness of developed/used systems. It is important to note that moral responsibility cannot be "contracted away," only liability can. Therefore, IAS providers and operators need to establish means of governance at the level of explicit morality. RRIAS provides justifications to establish, for example, accountability mechanisms. Vance et al. (2015) define (the organizational form of) accountability as the process of explaining one's action to another party who "has the right to pass judgment on those actions and to administer potential positive or negative consequences in response to them" (Vance et al. 2015, p. 347). These accountability mechanisms are a key element of many relationships between employees and managers (Vance et al. 2013) and thus also a key form of organisational governance. We understand the distribution of responsibility between provider and operator as both a requirement as well as a normative justification to establish such mechanisms and processes of accountability (Ananny and Crawford 2016; Martin 2019b).

Research on how algorithms support human-in-the-loop configurations (Grønsund and Aanestad 2020) and the triple-loop design approach (Seidel et al. 2018) highlight the important role of design for the ethical ramifications of IAS. Our egalitarian model proposes a continuum between IAS providers and IAS operators: The more providers grant control to operators, the more responsibility the operators will have. Furthermore, biases as a central issue of increasing algorithmic and autonomous system use (Kordzadeh and Ghasemaghaei 2021) can be minimized by adhering to RRIAS. Design choices particularly affect IAS providers in high-stake scenarios where potential harms of IAS use are more severe. We hold that all design processes

should consider the resulting design's ethical ramifications, especially for IAS. These general responsibilities also extend to how IAS providers communicate with operators via system interfaces. The providers should inform operators of the IAS's underlying functionality and scope such that the operators can make informed choices about whether an IAS with a particular reason-responsiveness is appropriate for from their planned usage, and if so, how they wish to implement proximal reason-responsiveness. This will impact retail and marketing processes, which consider the ethical ramifications of IAS. IAS operators are also directly targeted by RRIAS. They can ensure more streamlined processes if they decide to use IAS that are more reason-responsive. Our theory also shows that firms should strive to work closely with IAS providers to reduce responsibility gaps. We furthermore hold that IAS operators should lead the charge in taking responsibility for the IAS they use and for the responsibility issues those IAS pose.

Due to the normative orientation of our theory, it is particularly suitable to inform regulatory and judicial practices. Professional codes of ethics for IAS providers should consider RRIAS and the responsibilities of IAS providers and operators as guidance on moral responsibility. Lawmakers can be guided by the insights that large amounts of responsibility lie with (or are greatly influenced by) the IAS providers, and that regulations on the development of IAS might be of the most practical value. Notably, these responsibilities are moral responsibilities, not legal ones. Laws, consequently, could use these moral responsibilities as building blocks to be transformed into law. Based on RRIAS, it should rather be IAS operators than IAS providers who could be held accountable to use IAS that are reason-responsive and that enable employees to provide their proximal reasons to the IAS.

5 Limitations

First, since RRIAS is normative in nature and an ethical theory on the third level of normativity, it cannot directly be empirically verified or falsified; only counterarguments or impracticality can surpass it. Furthermore, our theory is limited due to its specific applicability to IAS as standard software. While RRIAS is able to cover IAS with different levels of partial autonomy, we stretch the borders of the applicability when it comes to IAS that have either no autonomy or that are fully autonomous. And despite recent advancements in artificial intelligence, a major share of IS used in firms nowadays fall in the category of IAS, therefore deploying a large impact across industries. Second, RRIAS is also limited based on its use case of standard software, expressed by the relationship between provider and operator. Other kinds of operator-provider relationships, especially more involved forms, are thus not covered by it. Further research should try to expand and/or adapt RRIAS to other use cases such as in-house development and more tight-knight operator provider relationships. Third, by distributing responsibility at a collective level, the individual responsibility of actors within these two firms is not explicitly covered. Future research should therefore expand our theory to other business relationships (e.g., freelancers, outsourcing) and responsibility at the individual sublevel, for example, settings where the IAS operator is involved in the development of the IAS, where IAS provider and operator collectives work within the same firm, or where specific employees misuse or inadequately design the system.

Funding Open access funding provided by University of Bern.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12599-023-00789-9.

References

Ananny M, Crawford K (2016) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 20(3):973–989. https://doi.org/10.1177/1461444816676645

Beck R, Dibbern J, Wiener M (2022) A multi-perspective framework for research on (sustainable) autonomous systems. Bus Inf Syst Eng 64(3):265–273. https://doi.org/10.1007/s12599-022-00752-0

Braham M, VanHees M (2011) Responsibility voids. Philos Q 61(242):6–15. https://doi.org/10.1111/j.1467-9213.2010.677.x

Collins S (2019) Collective responsibility gaps. J Bus Ethics 154(4):943–954. https://doi.org/10.1007/s10551-018-3890-6

Faraj S, Pachidi S, Sayegh K (2018) Working and organizing in the age of the learning algorithm. Inf Organ 28(1):62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Fischer JM, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. Cambridge University Press, Cambridge

French PA (1984) Collective and corporate responsibility. Columbia University Press, New York

Galliers RD, Newell S, Shanks G, Topi H (2017) Datification and its human, organizational and societal effects: the strategic opportunities and challenges of algorithmic decision-making. J Strateg Inf Syst 26(3):185–190. https://doi.org/10.1016/j.jsis.2017.08.002

Grønsund T, Aanestad M (2020) Augmenting the algorithm: emerging human-in-the-loop work configurations. J Strateg Inf Syst 29(2):101614. https://doi.org/10.1016/j.jsis.2020.101614

Harrington SJ (1996) The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. MIS Q 20(3):257–278. https://doi.org/10.2307/249656

Hellström T (2013) On the moral responsibility of military robots. Ethics Inf Technol 15(2):99–107. https://doi.org/10.1007/s10676-012-9301-2

Janiesch C, Fischer M, Winkelmann A, Nentwich V (2019) Specifying autonomy in the internet of things: the autonomy model and notation. Inf Syst e Bus Manag 17(1):159–194. https://doi.org/10.1007/s10257-018-0379-x

Jöhnk J, Weißert M, Wyrtki K (2021) Ready or not, AI comes. an interview study of organizational AI readiness factors. Bus Inf Syst Eng 63(1):5–20. https://doi.org/10.1007/s12599-020-00676-7

Johnson DG (2006) Computer systems: moral entities but not moral agents. Ethics Inf Technol 8(4):195–204. https://doi.org/10.1007/s10676-006-9111-5

Johnson DG (2015) Technology with no human responsibility? J Bus Ethics 127(4):707–715. https://doi.org/10.1007/s10551-014-2180-1

Kellogg KC, Valentine MA, Christin A (2019) Algorithms at work: the new contested terrain of control. Acad Manag Ann 14(1):366–410. https://doi.org/10.5465/annals.2018.0174

Khalil OEM (1993) Artificial decision-making and artificial ethics: a management concern. J Bus Ethics 12(4):313–321. https://doi.org/10.1007/BF01666535

Köchling A, Riazy S, Wehner MC, Simbeck K (2021) Highly accurate, but still discriminatory. Bus Inf Syst Eng 63(1):39–54. https://doi.org/10.1007/s12599-020-00673-w

Kordzadeh N, Ghasemaghaei M (2021) Algorithmic bias: review, synthesis, and future research directions. Eur J Inf Syst. https://doi.org/10.1080/0960085X.2021.1927212

Krenzer A, Stein N, Griebel M, Flath C (2019) Augmented intelligence for quality control of manual assembly processes using industrial wearable systems. In: ICIS 2019 Proceedings. https://aisel.aisnet.org/icis2019/mobile_iot/mobile_iot/9

Levy N (2017) The good, the bad, and the blameworthy. J Ethics Soc Philos 1(2):1–16. https://doi.org/10.26556/jesp.v1i2.6

Lüthi N, Matt C, Myrach T (2021) A value-sensitive design approach to minimize value tensions in software-based risk-assessment instruments. J Decis Syst 30(2–3):194–214. https://doi.org/10.1080/12460125.2020.1859744

Marabelli M, Newell S, Handunge V (2021) The lifecycle of algorithmic decision-making systems: organizational choices and ethical challenges. J Strateg Inf Syst 30(3):101683. https://doi.org/10.1016/j.jsis.2021.101683

Martin K (2019b) Ethical implications and accountability of algorithms. J Bus Ethics 160(4):835–850. https://doi.org/10.1007/s10551-018-3921-3

Martin K (2019a) Designing ethical algorithms. MIS Q Exec 18(2):Article 5. https://aisel.aisnet.org/misqe/vol18/iss2/5/

Mason E (2015) Moral ignorance and blameworthiness. Philos Stud 172(11):3037–3057. https://doi.org/10.1007/s11098-015-0456-7

Mason E (2019) Ways to be blameworthy: rightness, wrongness, and responsibility. Oxford University Press, Oxford

Matthias A (2004) The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics Inf Technol 6(3):175–183. https://doi.org/10.1007/s10676-004-3422-1

Mecacci G, Santoni de Sio F (2020) Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. Ethics Inf Technol 22(2):103–115. https://doi.org/10.1007/s10676-019-09519-w

Newell S, Marabelli M (2015) Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of 'datification.' J Strateg Inf Syst 24(1):3–14. https://doi.org/10.1016/j.jsis.2015.02.001

Pettit P (2007) Responsibility incorporated. Ethics 117(2):171–201. https://doi.org/10.1086/510695

Rivard S (2014) Editor's comments: the ions of theory construction. MIS Q 38(2):iii–xiv. https://www.jstor.org/stable/26634928

Rohner P (2013) Identity management for health professionals. Bus Inf Syst Eng 5(1):17–33. https://doi.org/10.1007/s12599-012-0244-2

Santoni de Sio F, Mecacci G (2021) Four responsibility gaps with artificial intelligence: why they matter and how to address them. Philos Technol. https://doi.org/10.1007/s13347-021-00450-x

Santoni de Sio F, van den Hoven J (2018) Meaningful human control over autonomous systems: a philosophical account. Front Robot AI 5:15. https://doi.org/10.3389/frobt.2018.00015

Sartorio C (2016) Causation and free will. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780198746799.001.0001

Seidel S, Berente N, Lindberg A, Lyytinen K, Nickerson JV (2018) Autonomous tools and design: a triple-loop approach to human-machine learning. Commun ACM 62(1):50–57. https://doi.org/10.1145/3210753

Sheridan TB, Parasuraman R (2005) Human-automation Interaction. Rev Hum Factors Ergon 1(1):89–129. https://doi.org/10.1518/155723405783703082

Sparrow R (2007) Killer Robots. J Appl Philos 24(1):62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Stahl B (2012) Morality, ethics, and reflection: a categorization of normative IS research. J Assoc Inform Syst 13(8):636–656. https://doi.org/10.17705/1jais.00304

Stahl BC (2006) Accountability and reflective responsibility in information systems. In: Zielinski C et al (eds) The information society: emerging landscapes. Springer, Boston, pp 51–68

Stephanidis C et al (2019) Seven HCI grand challenges. Int J Hum Comput Interact 35(14):1229–1269. https://doi.org/10.1080/10447318.2019.1619259

Vance A, Lowry PB, Eggett D (2013) Using accountability to reduce access policy violations in information systems. J Manag Inf Syst 29(4):263–290. https://doi.org/10.2753/MIS0742-1222290410

Vance A, Lowry PB, Eggett D (2015) Increasing accountability through user-interface design artifacts: a new approach to addressing the problem of access-policy violations. MIS Q 39(2):345–366. https://doi.org/10.25300/MISQ/2015/39.2.04

Wallach W, Allen C (2008) Moral machines: teaching robots right from wrong. Oxford University Press, Oxford

Zhou L, Paul S, Demirkan H, Yuan L, Spohrer J, Zhou M, Basu J (2021) Intelligence augmentation: towards building human-machine symbiotic relationship. AIS Transact Hum Comput Interact 13(2):243–264. https://doi.org/10.17705/1thci.00149