International Journal of Allied Medical Sciences and Clinical Research (IJAMSCR)

IJPAR |Vol.14 | Issue 4 | Oct - Dec -2025

www.ijamscr.com

DOI : https://doi.org/10.61096/ijpar.v14.iss4.2025.711 -723

ISSN: 2347-6567

Research

SELF AUDIT PROCEDURES FOR REGULATORY COMPLIANCE ISSUES

Gundavelly Akshith Kumar[1*], B. Sandeep[2], Dr. l. Harikiran[2]

[1,2] *Department Of Pharmaceutical Regulatory Affairs, Princeton College Of Pharmacy In Narapally, Ghatkesar, Telangana.*

[*]Author for Correspondence: Gundavelly Akshith Kumar
Email: princeton.pharmacy@gmail.com

Published on:

Published by:
Futuristic Publications

Abstract

In the past decade we have seen phenomenal improvements in the New Product and Service Introduction processes (NPSI). Applications of tools and methods, and use of process and quality management approaches have enabled companies to reduce continuously their cycle time, introduce products faster, cheaper, with higher quality, and better satisfy market needs. Among the enabling technologies, audits and assessments have played a critical role in driving these improvements. In major companies, these have become a permanent fixture of the NPSI processes. This paper discusses evolution of audits and assessments, and application of these tools from products to processes. The author presents recent developments in process audits for software and hardware development, some results achieved, and future trends.

Keywords: Audit, Assessment, Process Audit, Process Management, Quality, Software Quality

INTRODUCTION

Automated decision-making systems (ADMS), i.e. autonomous self-learning sys- tems that gather and process data to make qualitative judgements with little or no human intervention, increasingly permeate all aspects of society (AlgorithmWatch, 2019). This means that many decisions with significant implications for people and their environments—which were previously made by human experts—are now made by ADMS (Karanasiou & Pinotsis, 2017; Krafft et al., 2020; Zarsky, 2016). Exam- ples of the use of ADMS by both governments and private entities include poten- tially sensitive areas like medical diagnostics (Grote & Berens, 2020), recruitment (Sánchez-Monedero et al., 2020), driving autonomous vehicles (Evans et al., 2020), and the issuing of loans and credit cards

(Aggarwal et al., 2019; Lee et al., 2020). As information societies mature, the range of decisions that can be automated in this fashion will increase, and ADMS will be used to make ever-more critical decisions.

From a technical perspective, the specific models used by ADMS vary from sim- ple decision trees to deep neural networks (Lepri et al., 2018). In this paper, how- ever, we focus not on the underlying technologies but rather on the common features of ADMS from which ethical challenges arise. In particular, it is the combination of relative autonomy, complexity, and scalability that underpin both beneficial and problematic uses of ADMS (more on this in section Automated Decision-Making Systems). Delegating tasks to ADMS can help increase consistency, improve effi- ciency, and enable new solutions to complex problems (Taddeo & Floridi, 2018). Yet these improvements are coupled with ethical challenges. As noted already by Wiener (1988 [1954]): "The machine, which can learn and can make decisions on the basis of its learning, will in no way be obliged to make such decisions as we should have made, or will be acceptable to us." For example, ADMS may leave decision subjects vulnerable to harms associated with poor-quality outcomes, bias and discrimination, and invasion of privacy (Leslie, 2019). More generally, ADMS risk enabling human wrongdoing, reducing human control, removing human respon- sibility, devaluing human skills, and eroding human self-determination (Tsamados et al., 2020).

If these ethical challenges are not sufficiently addressed, a lack of public trust in ADMS may hamper the adoption of such systems which, in turn, would lead to sig- nificant social opportunity costs through the underuse of available and well-designed technologies (Cookson, 2018). Addressing the ethical challenges posed by ADMS is therefore becoming a prerequisite for good governance in information societies (Cath et al., 2018). Unfortunately, traditional governance mechanisms designed to oversee human decision-making processes often fail when applied to ADMS (Kroll et al., 2016). One important reason for this is that the delegation of tasks to ADMS curtails the sphere of ethical deliberation in decision-making processes (D'Agostino & Durante, 2018). In practice, this means that norms that used to be open for inter- pretation by human decision-makers are now embodied in ADMS. From an ethical

perspective, this shifts the focus of ethical deliberation from specific decision-mak- ing situations to the ways in which ADMS are designed and deployed.

From Principles to Practice

In response to the growing need to design and deploy ADMS in ways that are ethical, over 75 organisations—including governments, companies, academic institutions, and NGOs—have produced documents defining high-level guidelines (Jobin et al., 2019). Reputable contributions include Ethically Aligned Design (IEEE, 2019), Ethics Guidelines for Trustworthy AI (AI HLEG, 2019), and the OECD's Recommendation of the Council on Artificial Intelligence (OECD, 2019). Although varying in terminology, the different guidelines broadly con- verge around five principles: beneficence, non-maleficence, autonomy, justice, and explicability (Floridi & Cowls, 2019).

While a useful starting point, these principles tend to generate interpretations that are either too semantically strict, which are likely to make ADMS overly mechanical, or too flexible to provide practical guidance (Arvan, 2018). This indeterminacy hinders the translation of ethics principles into practices and leaves room for unethical behaviours like 'ethics shopping', i.e. mixing and matching ethical principles from different sources to justify some pre-existing behaviour; 'ethics bluewashing', i.e. making unsubstantiated claims about ADMS to appear more ethical than one is; and 'ethics lobbying', i.e., exploiting ethics to delay or avoid good and necessary legislation (Floridi, 2019). Moreover, the adoption of ethics guidelines remains voluntary, and the industry lacks both incentives and useful tools to translate principles into verifiable criteria (Raji et al., 2020). For example, interviews with software developers indicate that while they consider ethics important in principle, they also view it as an impractical construct that is distant from the issues they face in daily work (Vakkuri et al., 2019). Further, even organisations that are aware of the risks posed by ADMS may struggle to manage these, either due to a lack of useful governance mechanisms or conflict- ing interests (PwC, 2019). Taken together, there still exists a gap between the 'what' (and 'why') of ethics principles, and the 'how' of designing, deploying, and governing ADMS in practice (Morley et al., 2020).

A vast range of governance mechanisms that aim to support the translation of high-level ethics principles into practical guidance has been proposed in the existing literature. Some of these governance mechanisms focus on interventions in the early stages of software development processes, e.g. by raising the aware- ness of ethical issues among software developers (Floridi et al., 2018), creating more diverse teams of software developers (Sánchez-Monedero et al., 2020), embedding ethical values into technological artefacts through proactive design (Aizenberg and van den Hoven 2020; van de Poel, 2020), screening potentially biased input data (AIEIG, 2020), or verifying the

underlying decision-making models and code (Dennis et al., 2016). Other proposed governance mechanisms, such as impact assessments (ECP, 2018), take the outputs of ADMS into account. Yet others focus on the context in which ADMS operate. For example, so-called

Human-in-the-Loop protocols imply that human operators can either intervene to prevent or be held responsible for harmful system outputs (Jotterand & Bosco, 2020; Rahwan, 2018).

Scope, Limitations, and Outline

One governance mechanism that merits further examination is ethics-based audit- ing (EBA) (Diakopoulos, 2015; Raji et al., 2020; Brown et al., 2021; Mökander & Floridi, 2021). Operationally, EBA is characterised by a structured process whereby an entity's present or past behaviour is assessed for consistency with relevant princi- ples or norms (Brundage et al., 2020). The main idea underpinning EBA is that the causal chain behind decisions made by ADMS can be revealed by improved proce- dural transparency and regularity, which, in turn, allow stakeholders to identify who should be held accountable for potential ethical harms. Importantly, however, EBA does not attempt to codify ethics. Rather, it helps identify, visualise, and commu- nicate whichever normative values are embedded in a system. The aim thereby is to spark ethical deliberation amongst software developers and managers in organi- sations that design and deploy ADMS. This implies that while EBA can provide useful and relevant information, it does not tell human decision-makers how to act on that information. That said, by strengthening trust between different stakeholders and promoting transparency, EBA can facilitate morally good actions (more on this in section Ethics-based Auditing).

The idea of auditing software is not new. Since the 1970s, computer scientists have been researching how to ensure that different software systems adhere to pre- defined functionality and reliability standards (Weiss, 1980). Nor is the idea of auditing ADMS for consistency with ethics principles new. In 2014, Sandvig et al. referred to 'auditing of algorithms' as a promising, yet underexplored, governance mechanism to address the ethical challenges posed by ADMS. Since then, EBA has attracted much attention from policymakers, researchers, and industry practition- ers alike. For example, regulators like the UK Information Commissioner's Office (ICO) have drafted AI auditing frameworks (ICO, 2020; Kazim et al., 2021). At the same time, traditional accounting firms, including PwC (2019) and Deloitte (2020), technology-based startups like ORCAA (2020), and all-volunteer organisations like ForHumanity (2021) are all developing tools to help clients verify claims about their ADMS. However, despite a growing interest in EBA from both policymak- ers and private companies, important aspects of EBA are yet to be substantiated by academic research. In particular, a theoretical foundation for explaining how EBA affords good governance has hitherto been lacking.

In this article, we attempt to close this knowledge gap by analysing the feasibility and efficacy of EBA as a governance mechanism that allows organisations to opera- tionalise their ethical commitments and validate claims made about their ADMS. Potentially, EBA can also serve the purpose of helping individuals understand how a specific decision was made as well as how to contest it. Our primary focus, how- ever, is on the affordances and constraints of EBA as an organisational governance mechanism. The purpose thereby is to contribute to an improved understanding of what EBA is and how it can help organisations develop and deploy ethically-sound ADMS in practice.

To narrow down the scope of our analysis, we introduced two further limitations. First, we do not address any legal aspects of auditing. Rather, our focus in this arti- cle is on ethical alignment, i.e. on what ought and ought not to be done over and above compliance with existing regulation. This is not to say that hard governance mechanisms (like laws and regulations) are superfluous. In contrast, as stipulated by the AI HLEG (2019), ADMS should be lawful, ethical, and technically robust. How- ever, hard and soft governance mechanisms often complement each other, and deci- sions made by ADMS can be ethically problematic and deserving of scrutiny even when not illegal (Floridi, 2018). Hence, from now on, 'EBA' is to be understood as a soft yet formal1 'post-compliance' governance mechanism.

Second, any review of normative ethics frameworks remains outside the scope of this article. When designing and operating ADMS, tensions may arise between different ethical principles for which there are no fixed solutions (Kleinberg et al., 2017). For example, a particular ADMS may improve the overall accuracy of deci- sions but discriminate against specific subgroups in the population (Whittlestone et al., 2019a). Similarly, different definitions of fairness—like individual fairness, demographic parity, and equality of opportunity—are mutually exclusive (Friedler et al., 2016; Kusner et al., 2017). In short, it would be naïve to suppose that we have to (or indeed even

can) resolve disagreements in moral and political philosophy (see e.g. Binns, 2018) before we start to design and deploy ADMS. To overcome this challenge, we conceptualise EBA as a governance mechanism that can help organisations adhere to any predefined set of (coherent and justifiable) ethics prin- ciples (more on this in section Conceptual Constraints). EBA can, for example, take place within one of the ethical frameworks already mentioned, especially the Eth- ics Guidelines for Trustworthy AI (AI HLEG, 2019) for countries belonging to the European Union and the Recommendation of the Council on Artificial Intelligence (OECD, 2019) for countries that officially adopted the OECD principles. But organ- isations that design and deploy ADMS may also formulate their own sets of ethics principles and use these as a baseline to audit. The main takeaway here is that EBA is not morally good in itself, nor it is sufficient to guarantee morally good outcomes. EBA enables moral goodness to be realised, if properly implemented and combined with justifiable values and sincere intentions (Floridi, 2017a; Taddeo, 2016).

The remainder of this article proceeds as follows. In section Automated Decision- Making Systems, we define 'ADMS' and discuss the central features of ADMS that give rise to ethical challenges. In section Ethics-based Auditing, we explain what EBA is (or should be) in the context of ADMS. In doing so, we also clarify the roles and responsibilities of different stakeholders in relation to the EBA procedures. In section Status Quo: Existing EBA Frameworks and Tools, we provide an overview of currently available frameworks and tools for EBA of ADMS and how are these being implemented. We then offer three main contributions to the existing literature.

Status Quo: Existing EBA Frameworks and Tools

In this section, we survey the landscape of currently available EBA frameworks and tools. In doing so, we illustrate how EBA can provide new ways of detecting, under- standing, and mitigating the unwanted consequences of ADMS.

Ethics-based Auditing Frameworks

As described in the previous section, EBA frameworks are protocols that describe a specific EBA procedure and define what is to be audited, by whom, and according to which standards. Typically, EBA frameworks originate from one of two processes. The first type consists of 'top-down' national or regional strategies, like those pub- lished by the Government of Australia (Dawson et al., 2019) or Smart Dubai (2019). These strategies tend to focus on legal aspects or stipulate normative guidelines.2

At a European level, the debate was shaped by the AI4People project, which proposed that 'auditing mechanisms' should be developed to identify unwanted ethical consequences of ADMS (Floridi et al., 2018). Since then, the AI HLEG3 has published not only the Ethics-Guidelines for Trustworthy AI (2019), but also a corresponding Assessment List for Trustworthy AI (2020). This assessment list is intended for self-evaluation purposes and can thus be incorporated into EBA proce- dures. Such checklists are simple tools that help designers get a more informed view of edge cases and system failures (Raji et al., 2020). Most recently, the European Commission (2021) published its long-anticipated proposal of the new EU Artifi- cial Intelligence Act. The proposed regulation takes a risk-based approach. For our purposes, this means that a specific ADMS can be classified into one of four risk levels. While ADMS that pose 'unacceptable risk' are proposed to be completely banned, so-called 'high-risk' systems will be required to undergo legally mandated ex-ante and ex-post conformity assessments. However, even for ADMS that pose 'minimal' or 'limited' risk, the European Commission encourages organisations that design and deploy such systems to adhere to voluntary codes of conduct. In short, with respect to the proposed European regulation, there is a scope for EBA to help both providers of ADMS that pose limited risk to meet basic transparency obliga- tions and providers of high-risk systems to demonstrate adherence to organisational values that goes over and above what is legally required.

The second type of EBA frameworks emerges 'bottom-up', from the expansion of data regulation authorities to account for the effects ADMS have on informational privacy. Building on an extensive experience of translating ethical principles into governance protocols, frameworks developed by data regulation agencies provide valuable blueprints for EBA of ADMS. The CNIL privacy impact assessment, for example, requires organisations to describe the context of the data processing under consideration when analysing how well procedures align with fundamental ethical principles (CNIL, 2019). This need for contextualisation applies not only to data management but also to the use of ADMS at large. Another transferable lesson is that organisations should conduct an independent ethical evaluation of software they procure from—or outsource production to—third-party vendors (ICO, 2018). At the same time, EBA frameworks with roots in data regulation tend to account only for specific ethical concerns, e.g. those related to privacy. This calls for caution. Since there is a plurality of ethical values which may serve as

legitimate normative ends (think of freedom, equality, justice, proportionality, etc.), an exclusive focus on one, or even a few, ethical challenges risks leading to sub-optimisation from a holistic perspective.

To synthesise, the reviewed EBA frameworks converge around a procedure based on impact assessments. IAF (2019) summarised this procedure in eight steps: (1) Describe the purpose of the ADMS; (2) Define the standards or verifiable criteria based on which the ADMS should be assessed; (3) Disclose the process, including a full account of the data use and parties involved; (4) Assess the impact the ADMS has on individuals, communities, and its environment; (5) Evaluate whether the ben- efits and mitigated risks justify the use of ADMS; (6) Determine the extent to which the system is reliable, safe, and transparent; (7) Document the results and considera- tions; and (8) Reflect and evaluate periodically, i.e. create a feedback loop.

Ethics-based Auditing Tools

EBA tools are conceptual models or software products that help measure, evaluate, or visualise one or more properties of ADMS. With the aim to enable and facilitate EBA of ADMS, a great variety of such tools have already been developed by both academic researchers and privately employed data scientists. While these tools typi- cally apply mathematical definitions of principles like fairness, accountability and transparency to measure and evaluate the ethical alignment of ADMS (Keyes et al., 2019), different tools help ensure the ethical alignment of ADMS in different ways. A full review of all the tools that organisations can employ during EBA procedures would be beyond the scope of this article. Nevertheless, in what follows, we provide some examples of different types of tools that help organisations design and develop ethically-sound ADMS.4

Some tools facilitate the audit process by visualising the outputs of ADMS. FAIRVIS, for example, is a visual analytics system that integrates a subgroup dis- covery technique, thereby informing normative discussions about group fairness (Cabrera et al., 2019). Another example is Fairlearn, an open-source toolkit that treats any ADMS as a black box. Fairlearn's interactive visualisation dashboard helps users compare the performance of different models (Microsoft, 2020). These tools are based on the idea that visualisation helps developers and auditors to create more equitable algorithmic systems.

Other tools improve the interpretability of complex ADMS by generating more straightforward rules that explain their predictions. For example, Shapley Addi- tive exPlanations, or SHAP, calculates the marginal contribution of relevant fea- tures underlying a model's prediction (Leslie, 2019). The explanations provided by such tools are useful, e.g. when determining whether protected features have unjustifiably contributed to a decision made by ADMS. However, such explana- tions also have important limitations. For example, tools that explain the contri- bution of features that have been intentionally used as decision inputs may not determine whether protected features have contributed unjustifiably to a decision through proxy variables.

Yet other tools help convey the reasoning behind ADMS by applying one of three strategies: Data-based explanations provide evidence of a model by using compari- sons with other examples to justify decisions; Model-based explanations focus on the algorithmic basis of the system itself; and Purpose-based explanations focus on comparing the stated purpose of a system with the measured outcomes (Kroll, 2018). For our purposes, the key takeaway is that, while different types of expla- nations are possible, EBA should focus on local interpretability, i.e. explanations targeted at individual stakeholders—such as decision subjects or external auditors— and for specific purposes like internal governance, external reputation management, or third-party verification. Here, a parallel can be made to what Loi et al. (2020) call transparency as design publicity, whereby organisations that design or deploy ADMS are expected to publicise the intentional explanation of the use of a specific system as well as the procedural justification of the decision it takes.

Tools have also been developed that help to democratise the study of ADMS. Consider the TuringBox, which was developed as part of a time-limited research project at MIT. This platform allowed software developers to upload the source code of an ADMS so as to let others examine them (Epstein et al., 2018). The Turing- Box thereby provided an opportunity for developers to benchmark their system's

performance with regards to different properties. Simultaneously, the platform also allowed independent researchers to evaluate the outputs from ADMS, thereby add- ing an extra layer of procedural transparency to the software development process.

Finally, some tools help organisations document the software development pro- cess and monitor ADMS throughout their lifecycle. AI Fairness 360 developed by IBM, for example, includes metrics and algorithms to monitor, detect,

and mitigate bias in datasets and models (Bellamy et al., 2019). Other tools have been developed to aid developers in making pro-ethical design choices (Floridi, 2016b) by provid- ing useful information about the properties and limitations of ADMS. Such tools include end-user license agreements (Responsible AI Licenses, 2021), tools for detecting bias in datasets (Saleiro et al., 2018), and tools for improving transparency like datasheets for datasets (Gebru et al., 2018).

A Vision for Ethics-based Auditing of ADMS

Connecting the Dots

As demonstrated in section Status Quo: Existing EBA Frameworks and Tools above, a wide variety of EBA frameworks and tools have already been developed to help organisations and societies manage the ethical risks posed by ADMS. However, these tools are often employed in isolation. Hence, to be feasible and effective, EBA procedures need to combine existing conceptual frameworks and software tools into a structured process that monitors each stage of the software development lifecycle to identify and correct the points at which ethical failures (may) occur. In practice, this means that EBA procedures should combine elements of (a) functionality audit- ing, which focuses on the rationale behind decisions (and why they are made in the first place); (b) code auditing, which entails reviewing the source code of an algo- rithm; and (c) impact auditing, whereby the severity and prevalence of the effects of an algorithm's outputs are investigated.

It should be reemphasised that the primary responsibility for identifying and exe- cuting steps to ensure that ADMS are ethically sound rests with the management of the organisations that design and operate such systems. In contrast, the independent auditor's responsibility is to (i) assess and verify claims made by the auditee about its processes and ADMS and (ii) ensure that there is sufficient documentation to respond to potential inquiries from public authorities or individual decision subjects. More proactively, the process of EBA should also help spark and inform ethical deliberation throughout the software development process. The idea is that continu- ous monitoring and assessment ensures that a constant flow of feedback concerning the ethical behaviour of ADMS is worked into the next iteration of their design and application. Figure 2 below illustrates how the process of EBA runs in parallel with the software development lifecycle.

ADVANTAGES

- EBA of ADMS—as outlined in this article—displays six, interrelated and mutually reinforcing, methodological advantages. These are best illustrated by examples from existing tools:

- EBA can provide decision-making support to executives and legislators by defin- ing and monitoring outcomes, e.g. by showing the normative values embedded in a system (AIEIG, 2020). Here, EBA serves a diagnostic function: before ask- ing whether we would expect an ADMS to be ethical, we must consider which mechanisms we have to determine what it is doing at all. By gathering data on system states (both organisational and technical) and reporting on the same, EBA enables stakeholders to evaluate the reliability of ADMS in more detail. A systematic audit is thereby the first step to make informed model selection decisions and to understand the causes of adverse effects (Saleiro et al., 2018).

- EBA can increase public trust in technology and improve user satisfaction by enhancing operational consistency and procedural transparency. Mechanisms such as documentation and actionable explanations are essential to help indi- viduals understand why a decision was reached and contest undesired outcomes (Wachter et al., 2017). This also has economic implications. While there may be many justifiable reasons to abstain from using available technologies in certain contexts, fear and ignorance may lead societies to underuse available technolo- gies even in cases where they would do more good than harm (Cowls & Floridi, 2018). In such cases, increased public trust in ADMS could help unlock eco- nomic growth. However, to drive trust in ADMS, explanations need to be action- able and selective (Barredo Arrieta et al., 2020). This is possible even when algorithms are technically opaque since ADMS can be understood intentionally and in terms of their inputs and outputs.

- EBA allows for local alignment of ethics and legislation. While some normative metrics must be assumed when evaluating ADMS, EBA is a governance mecha- nism that allows organisations to choose which set of ethics principles they seek to adhere to. This allows for contextualisation. Returning to our example with fairness above, the most important aspect from an EBA perspective is not which specific definition of fairness is applied in a particular case, but that this deci- sion is communicated transparently and publicly justified. In short, by focusing on

identifying errors, tensions, and risks, as well as communicating the same to relevant stakeholders, such as customers or independent industry associations, EBA can help organisations demonstrate adherence to both sector-specific and geographically dependent norms and legislation.

• EBA can help relieve human suffering by anticipating potential negative conse- quences before they occur (Raji & Buolamwini, 2019). There are three overarch- ing strategies to mitigate harm: pre-processing, i.e. reweighing or modifying input data; in-processing, i.e. model selection or output constraints; and post- processing, i.e. calibrated odds or adjustment of classifications (Koshiyama, 2019). These strategies are not mutually exclusive. By combining minimum requirements on system performance with automated controls, EBA can help both developers test and improve the performance of ADMS (Mahajan et al., 2020) and enable organisations to establish safeguards against unexpected or unwanted behaviours.

• EBA can help balance conflicts of interest. A right to explanation must, for example, be reconciled with jurisprudence and counterbalanced with intellectual property law as well as freedom of expression (Wachter et al., 2017). By con- taining access to sensitive parts of the review process to authorised third-party auditors, EBA can provide a basis for accountability while preserving privacy and intellectual property rights.

• EBA can help human decision-makers to allocate accountability by tapping into existing internal and external governance structures (Bartosch et al., 2018). Within organisations, EBA can forge links between non-technical executives and developers. Externally, EBA help organisations validate the functionality of ADMS. In short, EBA can clarify the roles and responsibilities of different stakeholders and, by leveraging the capacity of institutions like national civil courts, help to redress the harms inflicted by ADMS.

• Naturally, the methodological advantages highlighted in this section are poten- tial and far from being guaranteed. However, the extent to which these benefits can be harnessed in practice depends not only on complex contextual factors but also on how EBA frameworks are designed. To realise its full potential as a governance mechanism, EBA of ADMS needs to meet specific criteria. In the next section, we turn to specifying these criteria.

Appendix B: Methodology

As mentioned in the introduction, the purpose of this article was to contribute to an improved understanding of what EBA is and how it can help organisations develop and deploy ethically-sound ADMS in practice. To achieve this aim, we let the fol- lowing three questions guide the research that led up to this article:

AIM AND OBJECTIVE

Self-inspection is a key part of the blood establishment quality management system and the base of different types of assessments. Self-inspection shall identify if there are problems, deficiencies or non-compliances against the quality policy, standard operating procedures (SOPs), guidelines, standards and regulations.

DISCUSSION:

Constraints Associated with Ethics-based Auditing

Criteria for Successful Implementation

Best practices for EBA of ADMS have yet to emerge. Nevertheless, as discussed in section Status Quo: Existing EBA Frameworks and Tools, organisations and researchers have already developed, and attempted to pilot, a wide range of EBA tools and frameworks. These early attempts hold valuable and generalisable lessons for organisations that wish to implement feasible and effective EBA procedures. As we will see, some of these lessons concern how stakeholders view EBA of ADMS, whilst other lessons concern the design of EBA practices. In this section, we will discuss the most important lessons from previous work and condense these into cri- teria for how to get EBA of ADMS right.

As a starting point, it should be acknowledged that ADMS are not isolated tech- nologies. Rather, ADMS are both shaped by and help shape larger sociotechnical systems. Hence, system output cannot be considered biased or erro-neous without some knowledge of the available alternatives. Holistic approaches to EBA of ADMS must therefore seek input from diverse stakeholders, e.g. for an inclusive discourse about key performance indicators (KPI). However, regardless of which KPI an organisation chooses to adopt, audits are only meaningful insofar as they

allow organisations to verify claims made about their ADMS. This implies that EBA procedures themselves must be traceable. By providing a traceable log of the steps taken in the design and development of ADMS, audit trails can help organisa- tions verify claims about their engineered systems. Here, a distinction should be made between traceability and transparency: while transpar- ency is often invoked to improve trust, full transpar- ency concerning the content of audits may not be desirable (e.g. with regards to privacy- and intellectual property rights). Instead, what counts is procedural trans- parency and regularity.

Further, to ensure that ADMS are ethically-sound, organisational policies need to be broken down into tasks for which individual agents can be held accountable. By formalising the software development process and revealing (parts of) the causal chain behind decisions made by ADMS, EBA helps clarify the roles and responsibilities of different stakeholders, including execu- tives, process owners, and data scientists. However, allocating responsibilities is not enough. Sustaining a culture of trust also requires that people who breach ethical and social norms are subject to proportional sanctions. By providing avenues for whistle-blowers and promoting a culture of ethical behaviour, EBA also helps strengthen interpersonal accountability within organisations. At the same time, doing the right thing should be made easy. This can be achieved through strategic governance structures that align profit with purpose. The 'trustworthiness' of a specific ADMS is never just a question about technology but also about value alignment. In practice, this means that the checks and balances developed to ensure safe and benevolent ADMS must be incorporated into organisational strategies, policies, and reward structures.

Importantly, EBA does not provide an answer sheet but a playbook. This means that EBA of ADMS should be viewed as a dialectic process wherein the auditor ensures that the right questions are asked and answered ade- quately. This means that auditors and systems owners should work together to develop context-specific methods. To manage the risk that independent auditors would be too easy on their clients, licences should be revoked from both auditors and system owners in cases where ADMS fail. However, it is dif- ficult to ensure that an ADMS contains no bias or to guarantee its fairness. Instead, the goal from an EBA perspective should be to provide useful information about when an ADMS is causing harm or when it is behaving in a way that is different from what is expected. This pragmatic insight implies that audits need to monitor and evaluate system outputs continuously, i.e. through 'oversight programs', and document performance characteristics in a comprehensible way. Hence, continuous EBA of ADMS implies considering system impacts as well as organisations, people, processes, and products.

Finally, the alignment between ADMS and specific ethical values is a design question. Ideally, properties like interpretability and robustness should be built into systems from the start, e.g. through 'Value-Aligned Design'. However, the context-dependent behaviour of ADMS makes it difficult to anticipate the impact ADMS will have on the complex environments in which they operate. By incorporating an active feedback element into the software development process, EBA can help inform the continuous re-design of ADMS. Although this may seem radical, it is already happening: most sciences, including engineering and jurisprudence, do not only study their systems, they simultaneously build and modify them.

Taken together, these generalisable lessons suggest that EBA procedures, even imperfectly implemented, can make a real difference to the ways in which ADMS are designed and deployed. However, our analysis of previous work also finds that, in order to be feasible and effective, EBA procedures must meet seven criteria. More specifically, to help organisations manage the ethical risks posed by ADMS, we argue that EBA procedures should be:

(1) Holistic, i.e. treat ADMS as an integrated component of larger sociotechnical contexts

(2) Traceable, i.e. assign responsibilities and document decisions to enable follow-up

(3) Accountable, i.e. help link unethical behaviours to proportional sanctions

(4) Strategic, i.e. align ethical values with policies, organisational strategies, and incentives

(5) Dialectic, i.e. view EBA as a constructive and collaborative process

(6) Continuous, i.e. identify, monitor, evaluate, and communicate system impacts over time

In addition to identifying the dimensions and indicators of TDC, we need to know the different levels of development of this competence, and to have valid instruments that allow educational institutions to grade this development in terms of learning, and guide future teachers to acquire their competencies through continuous improvement. Evaluating TDC in initial teacher training presents important challenges because of the inner

difficulties of evaluating competencies and establishing a framework for assessment. New tools are needed to help students reflect on situations and problems in line with the indicators to be evaluated. In recent years, various TDC evaluation tests have been developed on the basis of standards that make it possible to analyze the TDC level of teachers and future teachers:

a) The Wayfind Teacher Assessment mea- sures the use that teachers make of technology. This self-assessment test for teachers is already in use.

b) Selfie is a self-assessment tool based on self-perception, and its version for education centers and organizations has been imple- mented. It applies the European Commission's DigCompEdu as a reference standard. The rubric for the assessment has 6 levels of development ranging from newcomer to pioneer, which matches the model used for classifying linguis- tic competence. As the European Commission points out, it is a reference framework that must be contextualized.

c) The TDC Portfolio (INTEF, 2017) is an assessment system produced by the Spanish government for teachers based on the Common Framework standard of TDC. Teachers provide information about the assessment indicators. On the basis of this proposal, Tourón and colleagues (2018), developed an online self-assessment ques- tionnaire to determine the respondent's self-perception.

These are the reference frameworks of the instrument proposed in this paper. COMDID-A is a self-assessment tool aligned to the proposal made by the Catalan government (Generalitat de Catalunya, 2018), and to the Spanish and European contexts, as outlined by Lázaro and colleagues (2019).

This article aimed to study COMDID-A as a measure of the level of TDC among initial teacher training students. The PCA and Cronbach Alpha results are very good in terms of validity and show that this tool can now be applied to other samples to continue with the external validation. They also enable us to determine whether to include the items that do not have enough weight in any of the four factors found. These factors are the same as those defined by theory.

We shall now go on to discuss this part of the validity process per dimension. For D1. Teaching curriculum and methodology, all the items are within the theoretical dimension to which they were assigned in the processes of instrument creation and validation by experts. However, item

1.5: "When teaching, I include the guidelines of the educational institution for the integration of digital technologies in the classroom." also scores high on D2. This shows that student teachers believe that including these guidelines in the programming has to be planned. Currently, COMDID-A is being applied in several samples of in-service teachers in Catalonia to study whether this item should be accepted or not. In D2. Planning, organization and management, there are two items that were originally in D4. Item 4.5 "I train myself by doing activities related to digital technologies." seems to be understood as an organizational task rather than as an aspect of training in itself. In fact, the components of DT organization and management can be found as content in all teaching training activities. However, improving TDC in permanent teacher training has become one of the priorities for the professional development of practicing teachers. We believe that placing the item in D4, which deals with personal and professional development, is therefore justified. Item 4.3: "I use digital technologies with students, and I am a reference for using digital technology" seems to be understood by students as the organization and management of DT, rather than as a personal issue. This item is related to the role of leadership and being a model. Fundamental in the teaching profession, it is closely connected to personal abilities like communication, motivation, critical thinking, empathy, and personal safety. However, taking on the responsibility of being a reference or leader necessarily involves the ability to organize and manage digital technological resources, which implies being a source of inspiration for students (at the lowest level) and for colleagues (at a more advanced level), and is part of teachers' personal and professional development. Item 3.2. "I promote the access and use of digital technologies by all students with the intention of compensating for inequalities" on D2 second as well as D3. The first part of this item has more weight than the second. The guarantee of access to technology to all the students implicitly involves digital inclusion and the compensating function of education. Last but not least, item 2.4. "I follow the guidelines that schools prepare for teachers on the use of digital technologies in teaching" scored low on all dimensions. We believe that this item has to do with the relational part of the school, which is why the highest factor is D3. The item deals with the guidance and regulatory function of the documents of educational institutions, which are part of their educational project and autonomy. The scale D3. Relationship, ethics and security has the lowest reliability, but according to Hair et al. (2014) it is still very good. The first part of item 3.2 "I promote the access and use of digital

technologies by all students with the intention of compensating for inequalities" has more weight than the second part.

Access to technology must be guaranteed as part of the compensatory and regulatory function of inequalities that all schools and the education system in general must have. This is closely connected to D3. Finally, the scale D4. Personal and professional has the highest internal reliability. However, there is one item (3.5: "I access the contents distributed in different digital spaces of the educational center and comment on them [blogs, virtual environments, social networks, etc.])" that should be in D3 but in the PCA analysis is, in fact in, D4. We believe that this is because the personal part of social networks has more weight in this sentence. But when the instrument was constructed, this item was placed in D3 because the digital spaces of the educational center and their contents must be an institutional strategy (frequency of publication, recipients, objectives, communication strategy linked to projects, etc.). We argue that item 3.5 should remain in D3 because we agree with Marthese and Shu-Nu Chang (2017) that the responsible and ethical use of technology can be modeled, discussed and practiced by teachers but they need to be aware of DT and the new practices that they entail. Promoting the use of the digital spaces of the center should be part of an institutional.

communication and visibility strategy towards the outside. According to Marthese and Shu-Nu Chang (2017), the responsible and ethical use of technology can be modeled, discussed and practiced by teachers but they need to be aware of DT and the new practices that they entail. Therefore, this indicator should be in D3.

We will now move on to study the correlation of age, gender and university access with students' TDC in order to answer our study's second question. In our sample, age correlates significantly and negatively with D2, D4 and D3. The older the student, the lower the self-evaluation in three of the four dimensions, and average values are high. This may be related to the over-confidence with which future teachers approach DTs in general: younger people use technology in a more natural way, which is one of the characteristics of present-day students, and of self- evaluation of TD in particular. However, as reported by Roig and colleagues (2015), there is another aspect of age that should be taken into account: the number of years of teaching also correlates with factors of TD use. This contrasts with Prensky's (2001) postulates about digital natives, according to which young people tend to use TD more and better. In teaching practice, it seems clear that experience and age will determine the awareness and sureness with which teachers naturally appropriate and incorporate DTs into their daily activities. In agreement with, these results show that they have an acceptable level of basic TDC, but do not have an acceptable level of applying DT to teaching or of the digital strategies necessary for their own professional development. This also coincides with Cabero (2013), who concludes that teachers perceive that they are more than able to use DT in classrooms because they know several Office applications for work in class. However, they have little digital command of specific tools, for example, for designing online activities to complement or support teaching processes. Therefore, there is a need for specific training in these areas.

Although we have reported one particular tool for the self-assessment of TDC, the only way TDC can be measured and understood to be the complex process that it is will be to use a wide variety of different tools This study may help initial teacher training institutions by providing them with a valid and reliable assessment tool that complements the information from the curriculum and helps make proposals for reviewing and improving initial teacher training curricula. Furthermore, action can be taken to improve the training of future teachers on the basis of specific data obtained from the self-evaluation process of each dimension. The results show that initial teacher training should include strategies on how to manage information from the Internet, how to search for and select information, and how to communicate it to others. In order to accept the new roles and ways of teaching, future teachers must reflect and adopt the new models of teaching and learning.

CONCLUSION

In conclusion, self-inspection is part of a learning process, they should recognize the efforts given by the staff, will help to correct noncompliances effectively, and evaluate the facility's quality and operational systems to determine whether the service they provide is appropriate and in control.

ACKNOWLEDGEMENT

The Authors are thankful to the Management and Principal, Princeton College of Pharmacy, Narapally, Ghatkesar, Telangana, for extending support to carry out the research work. Finally, the authors express their gratitude to the Sura Pharma Labs, Dilsukhnagar, Hyderabad, for providing research equipment and facilities.

REFRENCES

1. Aggarwal, N., Eidenmüller, H., Enriques, L., Payne, J., & Zwieten, K. (2019). Autonomous systems and the law. München: Baden-Baden.

2. AI HLEG. 2019. European Commission's ethics guidelines for trustworthy artificial intelligence. https:// ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.

3. AIEIG. 2020. From principles to practice — An interdisciplinary framework to operationalise AI ethics. AI Ethics Impact Group, VDE Association for Electrical Electronic & Information Technologies e.V., Bertelsmann Stiftung, 1–56. https://doi.org/10.11586/2020013.

4. Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. Big Data and Society. https://doi.org/10.1177/2053951720949566

5. AlgorithmWatch. 2019. Automating society: Taking stock of automated decision-making in the EU. Bertelsmann Stiftung, 73–83. https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_ Society_Report_2019.pdf.

6. Ananny, M., & Crawford, K. (2018). Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media and Society, 20(3), 973–989. https://doi. org/10.1177/1461444816676645

7. Arvan, M. (2018). Mental time-travel, semantic flexibility, and A.I. ethics. AI and Society. https://doi.org/ 10.1007/s00146-018-0848-2

8. Assessment List for Trustworthy AI. 2020. Assessment list for trustworthy AI (ALTAI). https://ec.europa. eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-asses sment.

9. Auer, F., & Felderer, M. (2018). Shifting quality assurance of machine learning algorithms to live systems. In: M. Tichy, E. Bodden, M. Kuhrmann, S. Wagner, & J.-P. Steghöfer (Eds.), Software Engi- neering und Software Management 2018 (S. 211–212). Bonn: Gesellschaft für Informatik.

10. Barredo Arrieta, A., Del Ser, J., Gil-Lopez, S., Díaz-Rodríguez, N., Bennetot, A., Chatila, R., et al. (2020). Explainable explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j. inffus.2019.12.012

11. Bellamy, R. K. E., Mojsilovic, A., Nagar, S., Natesan Ramamurthy, K., Richards, J., Saha, D., Sattigeri, P., et al. (2019). AI fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM Journal of Research and Development. https://doi.org/10.1147/JRD.2019.2942287

12. Binns, R. (2018). What can political philosophy teach us about algorithmic fairness? IEEE Security & Privacy, 16(3), 73–80.

13. Boddington, P., Millican, P., & Wooldridge, M. (2017). Minds and machines special issue: Eth- ics and artificial intelligence. Minds and Machines, 27(4), 569–574. https://doi.org/10.1007/ s11023-017-9449-y

14. Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. Big Data & Society, 8(1), 205395172098386. https://doi.org/10.1177/2053951720983865

15. Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., et al. 2020. Toward trustworthy AI development: Mechanisms for supporting verifiable claims. ArXiv, no. 2004.07213[cs.CY]. http://arxiv.org/abs/2004.07213.

16. Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. Computer, 50(5), 116–19.

17. Burrell, Jenna. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms.

Big Data & Society. https://doi.org/10.1177/2053951715622512

18. Cabrera, Á. A., Epperson, W., Hohman, F., Kahng, M., Morgenstern, J., Chau, D. H. 2019. FairVis: Visual analytics for discovering intersectional bias in machine learning. http://arxiv.org/abs/1904. 05419.

19. Cath, C., Cowls, J., Taddeo, M., & Floridi, L. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A Math- ematical, Physical and Engineering Sciences. https://doi.org/10.1098/rsta.2018.0080

20. Chopra, A. K., Singh, M. P. 2018. Sociotechnical systems and ethics in the large. In AIES 2018—Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society (pp. 48–53). https://doi.org/ 10.1145/3278721.3278740.

21. Christian, B. (2020). The alignment problem: Machine learning and human values. W.W. Norton & Company Ltd.

22. Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. HeinOnline, 1, 1–34.

23. CNIL. 2019. Privacy impact assessment—Methodology. Commision Nationale Informatique & Libertés, 400.

24. Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. Science and Engineering Ethics, 26(4), 2051–2068. https://doi.org/10.1007/ s11948-019-00146-8

25. Conrad, C. A. (2018). Business ethics—A philosophical and behavioral approach. Springer. https://doi. org/10.1007/978-3-319-91575-3

26. Cookson, C. 2018. Artificial intelligence faces public backlash, warns scientist. Financial Times, June 9, 2018. https://www.ft.com/content/0b301152-b0f8-11e8-99ca-68cf89602132.

27. Council of Europe. 2018. Algorithms and human rights. www.coe.int/freedomofexpression.

28. Cowls, J., & Floridi, L. (2018). Prolegomena to a white paper on an ethical framework for a good AI society. SSRN Electronic Journal.

29. Cummings, M. L. 2004. Automation bias in intelligent time critical decision support systems. In Collection of technical papers—AIAA 1st intelligent systems technical conference (Vol. 2, pp. 557–62).

30. Dafoe, A. (2017). AI governance: A research agenda. American Journal of Psychiatry. https://doi.org/10. 1176/ajp.134.8.aj1348938

31. D'Agostino, M., & Durante, M. (2018). Introduction: The governance of algorithms. Philosophy and Technology, 31(4), 499–505. https://doi.org/10.1007/s13347-018-0337-z

32. Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., Scowcroft J, and Hajkowicz S. 2019. Artificial intelligence: Australia's ethics framework.

33. Deloitte. 2020. Deloitte introduces trustworthy AI framework to guide organizations in ethical applica- tion of technology. Press Release. 2020. https://www2.deloitte.com/us/en/pages/about-deloitte/ articles/press-releases/deloitte-introduces-trustworthy-ai-framework.html.

34. Dennis, L. A., Fisher, M., Lincoln, N. K., Lisitsa, A., & Veres, S. M. (2016). Practical verification of decision-making in agent-based autonomous systems. Automated Software Engineering, 23(3), 305–359. https://doi.org/10.1007/s10515-014-0168-9.

35. Di Maio, P. (2014). Towards a metamodel to support the joint optimization of socio technical systems. Systems, 2(3), 273–296. https://doi.org/10.3390/systems2030273

36. Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. Digital Journalism, 3(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

37. Dignum, V. 2017. Responsible autonomy. In Proceedings of the international joint conference on autonomous agents and multiagent systems, AAMAS 1: 5. https://doi.org/10.24963/ijcai.2017/655. ECP. 2018. Artificial intelligence impact assessment.

38. Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published From 1940 Through 2017. Personality and Social Psychology Review, 23(4), 332–366. https://doi.org/10.1177/1088868318811759

39. Epstein, Z., Payne, B. H., Shen, J. H., Hong, C. J., Felbo, B., Dubey, A., Groh, M., Obradovich, N., Cebrian, M., Rahwan, I. 2018. Turingbox: An experimental platform for the evaluation of AI sys- tems. In IJCAI international joint conference on artificial intelligence 2018-July (pp. 5826–28). https://doi.org/10.24963/ijcai.2018/851.

40. Erdelyi, O. J., Goldsmith, J. 2018. Regulating artificial intelligence P. In AAAI/ACM conference on artificial intelligence, ethics and society. http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_13.pdf.

41. Etzioni, A., & Etzioni, O. (2016). AI assisted ethics. Ethics and Information Technology, 18(2), 149–156. https://doi.org/10.1007/s10676-016-9400-6

42. European Commission. 2021. Proposal for regulation of the European Parliament and of the council. COM(2021) 206 final. Brussels.

43. Evans, K., de Moura, N., Chauvier, S., Chatila, R., & Dogan, E. (2020). Ethical decision making in autonomous vehicles: The AV ethics project. Science and Engineering Ethics, 26(6), 3285–3312. https://doi.org/10.1007/s11948-020-00272-8

44. Fagerholm, F., Guinea, A. S., Mäenpää, H., Münch, J. 2014. Building blocks for continuous experimentation. In Proceedings of the 1st international workshop on rapid continuous software engineering (pp. 26–35). RCoSE 2014. ACM. https://doi.org/10.1145/2593812.2593816.

45. Falkenberg, L., & Herremans, I. (1995). Ethical behaviours in organizations: Directed by the formal or informal systems? Journal of Business Ethics, 14(2), 133–143. https://doi.org/10.1007/BF008 72018

46. Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. Science and Engineering Ethics, 26(6), 3333–3361. https://doi. org/10.1007/s11948-020-00276-4

47. Floridi, L. (2013). Distributed morality in an information society. Science and Engineering Ethics, 19(3), 727–743. https://doi.org/10.1007/s11948-012-9413-4.

48. Floridi, L. (2016a). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083). https://doi.org/10.1098/rsta.2016.0112.

49. Floridi, L. (2016b). Tolerant paternalism: Pro-ethical design as a resolution of the dilemma of toleration.

50. Science and Engineering Ethics, 22(6), 1669–1688. https://doi.org/10.1007/s11948-015-9733-2. Floridi, L. (2017a). Infraethics–On the conditions of possibility of morality. Philosophy and Technology, 30(4), 391–394. https://doi.org/10.1007/s13347-017-0291-1.