Digital Society (2022) 1:21
https://doi.org/10.1007/s44206-022-00022-2

ORIGINAL PAPER

Continuous Auditing of Artificial Intelligence: a Conceptualization and Assessment of Tools and Frameworks

Matti Minkkinen[1] · **Joakim Laine**[1] · **Matti Mäntymäki**[1]

Received: 24 March 2022 / Accepted: 15 September 2022 / Published online: 4 October 2022

Abstract

Artificial intelligence (AI), which refers to both a research field and a set of technologies, is rapidly growing and has already spread to application areas ranging from policing to healthcare and transport. The increasing AI capabilities bring novel risks and potential harms to individuals and societies, which auditing of AI seeks to address. However, traditional periodic or cyclical auditing is challenged by the learning and adaptive nature of AI systems. Meanwhile, continuous auditing (CA) has been discussed since the 1980s but has not been explicitly connected to auditing of AI. In this paper, we connect the research on auditing of AI and CA to introduce CA of AI (CAAI). We define CAAI as a (nearly) real-time electronic support system for auditors that continuously and automatically audits an AI system to assess its consistency with relevant norms and standards. We adopt a bottom-up approach and investigate the CAAI tools and methods found in the academic and grey literature. The suitability of tools and methods for CA is assessed based on criteria derived from CA definitions. Our study findings indicate that few existing frameworks are directly suitable for CAAI and that many have limited scope within a particular sector or problem area. Hence, further work on CAAI frameworks is needed, and researchers can draw lessons from existing CA frameworks; however, this requires consideration of the scope of CAAI, the human–machine division of labour, and the emerging institutional landscape in AI governance. Our work also lays the foundation for continued research and practical applications within the field of CAAI.

Keywords Artificial intelligence · Continuous auditing · AI auditing · Continuous AI auditing · Algorithmic auditing

✉ Matti Minkkinen
matti.minkkinen@utu.fi

[1] Turku School of Economics, University of Turku, 20014 Turku, Finland

1 Introduction

Artificial intelligence (AI), which refers to both a research field and a set of technologies, is rapidly growing and has already spread to application areas ranging from policing to healthcare and transport (e.g. Rezende, 2020; Stilgoe, 2018; Trocin et al., 2021). The growth in AI applications is set to continue in the near term, and in the long term, AI technologies can transform areas such as scientific methods, foreign policy, and personalized medicine (Tewari, 2022). In general, AI is integrated into information systems and refers to the capabilities of data interpretation, learning, and adaptation that aim to attain human-level capabilities in particular tasks (Kaplan & Haenlein, 2019; Russell & Norvig, 2021). In some cases—for example the optimization of online search results and the filtering of social media feeds—AI has already become commonplace and near invisible.

The increasing AI capabilities and applications bring novel risks and potential harms for individuals and societies, such as lack of transparency and accountability, as well as biases against individuals and groups (Dignum, 2020; Floridi et al., 2018; Martin, 2019). These challenges and risks related to AI systems underscore the importance of AI governance at the organizational, interorganizational, and societal levels (Laato et al., 2022; Mäntymäki et al., 2022a, b; Minkkinen et al., 2022a, b; Schneider et al., 2022; Seppälä et al., 2021). As a closely related parallel to governance, auditing of AI is promoted as a means of tackling risks by holding AI systems and organizations that use AI to certain criteria and by requiring necessary controls (Koshiyama et al., 2021; Minkkinen et al., 2022a, b; Mökander et al., 2021; Sandvig et al., 2014). In addition to tackling risks, auditing of AI has been promoted as a new industry and a source of economic growth (Koshiyama et al., 2021). Nevertheless, auditing faces challenges owing to the nature of some AI technologies. Traditionally, auditing has been conducted periodically or cyclically, in which case audits represent snapshots of systems and processes. In snapshot audits, timing is crucial because an early audit can influence an AI system's design and operations more than a post-deployment audit of a production system can (Raji et al., 2020b; cf. Laato et al., 2022a, b). Whilst many AI systems use fairly static models with periodic updates, some systems, such as those based on reinforcement learning, adapt as a result of highly complex models, which means that they may exhibit unpredictable results (Dignum, 2020; Falco et al., 2021; Kaplan & Haenlein, 2019). Learning and adaptation present benefits but also potential risks, as AI systems learn patterns that are not hard-coded by designers. Adaptation presents a specific challenge for snapshot auditing because a system that is deemed compliant at one point may not be compliant later. In addition, the operating and evolution speeds of AI systems are much faster than those of human-led snapshot auditing processes, which are usually relatively cumbersome.

As the challenges of snapshot audits were already apparent before the recent growth of AI adoption, the continuous auditing (CA) concept was introduced in 1989 (Groomer & Murthy, 1989) in response to the need for near-real-time auditing

information. Auditing AI and CA are a natural match because CA can potentially keep pace with the AI system's evolution and continuously provide up-to-date information on its performance according to set criteria. The rationale for CA is linked to the aspired human oversight of AI systems (Floridi et al., 2018; Shneiderman, 2020). On the one hand, CA may challenge human agency by transferring part of auditing to machines, but on the other hand, it may also free human capacity to conduct higher-level auditing tasks. Provisionally, CA of AI systems appears most relevant to organizations' internal audit functions (cf. Raji et al., 2020b; Tronto et al., 2021) as opposed to external auditing conducted by independent auditors, although this may change as the audit ecosystem continues to evolve (Mökander et al., 2022).

The potential of continuous AI auditing approaches has already been noted by the European Union (EU), whose proposed that AI Act (European Commission, 2021) includes provisions for the mandatory post-market monitoring of high-risk AI systems. In the proposed EU regulation, the providers of high-risk AI systems would need to draft post-market monitoring plans to document the performance of these systems throughout their life cycles after they are introduced to the market (Mökander et al., 2022). However, although CA is a mature concept (e.g. Eulerich & Kalinichenko, 2018; Shiue et al., 2021; Vasarhelyi & Halper, 1991), we were unable to find an established literature stream specifically on CA of AI (CAAI) beyond general calls for monitoring the impacts of algorithmic systems (e.g. Doneda & Almeida, 2016; Metcalf et al., 2021; Shah, 2018; Yeung et al., 2020).

To address the paucity of the CAAI literature, this study has been positioned to answer the following research question: *What is continuous auditing of artificial intelligence, and what frameworks and tools exist for its execution?* The current paper advances the body of knowledge on auditing of AI (Brown et al., 2021; Koshiyama et al., 2021; Mökander et al., 2021; Sandvig et al., 2014) in two ways. First, we connect the research on auditing of AI and CA, introducing the CAAI concept. Second, we present an assessment of the suitability of AI auditing frameworks and tools for CAAI. In particular, we adopt a bottom-up approach and investigate tools and methods for CAAI. By conceptualizing CAAI and surveying frameworks and tools, this study lays the foundation for continued research and practical applications within the field of CAAI.

The remainder of the paper is structured as follows. In the Sect. 2, we introduce auditing of AI and CA and posit that CAAI lies at the intersection of the two. We then present our materials and methods, providing an overview of the examined auditing frameworks and tools and our assessment criteria for CA. The Sect. 4 assesses the suitability of the frameworks and tools for CA. The paper ends with the Sect. 5, which lays out the state of the art in CAAI, explores lessons from existing CA frameworks, and discusses limitations and future research directions.

2 Conceptual Background

2.1 Auditing of AI

The literature discusses auditing of AI under various terms. The early literature (Sandvig et al., 2014) and subsequent research (Brown et al., 2021; Galdon Clavell et al., 2020; Koshiyama et al., 2021) refer to "algorithm auditing" as a means to discover and mitigate discrimination and other problematic consequences of the use of algorithms. Interest in auditing algorithms has grown in conjunction with the increasing capabilities and power of inscrutable "black-box" algorithms that support decision-making and impact people and organizations (Pasquale, 2015).

The recent literature has introduced the concept of the ethics-based auditing (EBA) of automated decision-making systems (Mökander et al., 2021). EBA is defined as "a structured process whereby an entity's present or past behaviour is assessed for consistency with relevant principles or norms" (Mökander et al., 2021, p. 1). This definition usefully leaves the audited entity open; thus, the targets of auditing may be algorithms, AI systems, or organizations. Brown et al., (2021, p. 2), in turn, defined ethical algorithm audits as "assessments of the algorithm's negative impact on the rights and interests of stakeholders, with a corresponding identification of situations and/or features of the algorithm that give rise to these negative impacts". The difference between these two definitions is that ethical algorithm audits focus on impact, whilst EBA highlights consistency with principles and norms. The definition of the ethical algorithm audit (Brown et al., 2021) also posits algorithms as the target of auditing rather than leaving the audited entity open.

For our conceptualization and assessment, we consider auditing of AI to encompass both principle- and impact-based approaches, preferring not to delimit the field prematurely. We acknowledge the existence of several types of AI auditing, such as auditing system performance. For example, according to the EU High-Level Expert Group, trustworthy AI consists of three components: AI should be lawful, ethical, and technically robust (High-Level Expert Group on Artificial Intelligence, 2019). These components are not fully independent of each other; for example, ethical concerns can lead to legal consequences, and a lack of technical robustness can lead to ethical concerns (Floridi et al., 2022). However, in the following discussion of CAAI, our primary focus is on the consideration of ethical issues and potential harm, such as matters of safety, in line with most of the current literature on auditing of AI (e.g. Falco et al., 2021; High-Level Expert Group on Artificial Intelligence, 2019; Mökander et al., 2021). A further argument in favour of an ethics focus is that companies have economic incentives to develop high-performing AI systems, but auditing to ensure safe and ethically responsible AI requires further research on tools and frameworks.

We use the term Sect. 2.1 to highlight that our study focuses on auditing *of* AI rather than auditing using AI. There is a separate and growing stream of literature on the use of AI and other novel technologies to aid auditing (e.g. Kokina & Davenport, 2017). In contrast to this literature, we investigate auditing of AI systems to discover and mitigate potential risks, harms, and breaches of standards. Whilst technical tools

may play a significant role in auditing, in our study, AI is the target of auditing rather than the means.

2.2 Continuous Auditing

Traditionally, auditing procedures have been performed on a cyclical basis—for example once a month—after business activities have occurred (Coderre, 2005). Breaking with this cyclical approach, CA was first introduced by Groomer and Murthy (1989), and then Vasarhelyi and Halper (1991) applied a monitoring layer for auditors (Shiue et al., 2021; Yoon et al., 2021). Whilst the concept of CA has existed since the 1980s, and multiple definitions have been presented, no standard definition exists. The American Institute of Certified Public Accountants (AICPA, 1999) defined CA as "a methodology that enables independent auditors to provide written assurance on a subject matter using a series of auditors' reports issued simultaneously with, or a short time after, the occurrence of events underlying the subject matter". Focusing on the auditing component, the Institute of Internal Auditors defined internal auditing as follows:

an independent activity of objective assessment and of consulting designed to add value and improve operations of organizations while achieving their objectives through a systematic and disciplined approach in the evaluation of effectiveness of risk management, control and governance processes. (Institute of Internal Auditors, 2022).

Thus, auditing in general aims to serve organizations by evaluating risk management, controls, and governance, and CA introduces a further real-time component to it.

Compared to traditional auditing, CA features more frequent audits, a more proactive model, and automated procedures (Yoon et al., 2021). CA definitions include elements such as the processes of collecting and evaluating data, ensuring the real-time efficiency and effectiveness of systems, and performing controls and risk assessments automatically (Coderre, 2005; Marques & Santos, 2017). Two main activities emerge with CA: continuous control and risk assessments (Coderre, 2005). They focus on auditing systems as early as possible and highlight processes or systems that experience higher-than-expected levels of risk. In addition, CA changes the role of the auditor; the nature, timing, and extent of the auditing; and the nature of audit reporting, data modelling, data analytics, and monitoring (Yoon et al., 2021). In particular, the role of internal auditors has changed, as they not only control audit activities but also monitor risk controls and identify areas in which risk management processes can be improved (Coderre, 2005). Eulerich and Kalinichenko (2018, p. 33) synthesized previous definitions and defined CA as follows:

a (nearly) real-time electronic support system that continuously and automatically audits clearly defined "audit objects" based on pre-determined criteria. CA identifies exceptions and/or deviations from a defined standard or benchmark, and reports them to the auditor. With this continuous approach, the audit occurs within the shortest possible time after the occurrence of an event.

CA brings many benefits. It reduces risks, diminishes fraud attempts, facilitates the objectives of internal control, allows timely access to information, integrates internal and external stakeholders and helps external auditing, allows timely

adjustments, and modifies auditors' routine tasks, thereby allowing them to focus on more important responsibilities (Marques & Santos, 2017). Moreover, it increases confidence in transactions and operational processes, decision-making, and financial statements (Marques & Santos, 2017). Audit executives often prefer ongoing assessments rather than periodic reviews (Coderre, 2005). The next stage in audit development is CA utilizing computer science technologies, as researchers have provided solutions to the development of CA in organizational auditing (Wang et al., 2020).

2.3 Towards Continuous Auditing of Artificial Intelligence

Drawing on CA and auditing of AI, this study introduces the concept of CAAI, which is a type of auditing that exists at the intersection of CA and auditing of AI (Fig. 1). CAAI is CA that targets AI systems and corresponding organizations. In other words, CA provides the *auditing methods*, and auditing of AI provides the *audit object*. The intersectional position of CAAI means that it is a subset of both CA and auditing of AI. Not all CA targets AI systems; conversely, not all auditing of AI uses continuous approaches.

The following is our working definition of CAAI: *CAAI is a (nearly) real-time electronic support system for auditors that continuously and automatically audits an AI system to assess consistency with relevant norms and standards.*

In line with a recent definition (Eulerich & Kalinichenko, 2018), we conceptualize CA as a (nearly) real-time electronic support system for auditors. Because CA definitions emphasize the automated nature of auditing, we decided to delimit the concept to the technical component. Nevertheless, CA operates in socio-technical systems together with human auditors. The AI system is posited as the audit target, which gives CAAI a clear focus and differentiates it from other types of auditing, such as financial auditing. The investigated AI system gives boundaries to CAAI, and eventually, organizations may complement it with broader auditing practices. Moreover, we draw on the EBA definition to highlight consistency with particular norms and standards (Mökander et al., 2021). These are defined by law, ethics, and

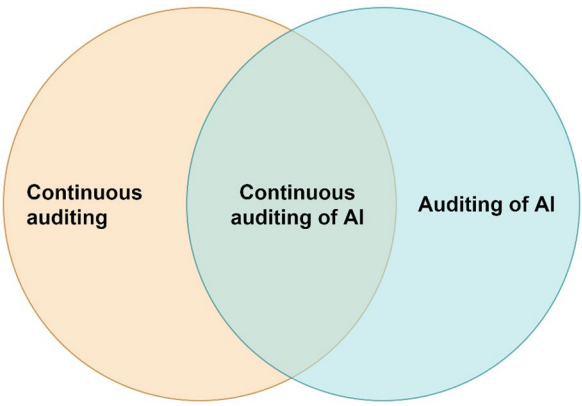

Fig. 1 Continuous auditing of artificial intelligence at the intersection of continuous auditing and auditing of artificial intelligence

societal norms, and they change over time. In the case of AI systems, the relevant norms and standards can also entail the examination of a system's potential impacts. Compared to the EBA definition (Mökander et al., 2021), we omit "principles" because, in our view, for principles to be continuously audited, they need to be operationalized into norms and standards.

Like continuous auditing generally, CAAI markedly changes the temporality and tempo of auditing, whereby the audit of past or present events becomes the almost real-time monitoring of current events. Hence, the temporality of auditing comes closer to that of audited AI systems. Because CAAI requires continuous access to AI systems, it appears most relevant to internal audit functions within organizations (cf. Raji et al., 2020b; Tronto et al., 2021) as opposed to external auditing conducted by independent auditors. However, internal and external auditing roles may develop as the audit ecosystem evolves (Mökander et al., 2022). CA also changes the division of labour between humans and machines because the auditor can focus their attention on more interpretive and complex tasks rather than on processing data (Eulerich & Kalinichenko, 2018).

3 Materials and Methods

3.1 Overview of Studied Papers

In this study, we assessed AI auditing tools and frameworks vis-à-vis their suitability to CAAI. Table 1 presents the descriptive details of the papers included in this assessment. The papers were selected based on targeted searches of auditing together with AI or near-synonyms, such as "machine learning", "deep learning", "algorithm", and "black box". The goal was to summarize the most important AI auditing tools and frameworks and review their suitability for CAAI. We selected studies that addressed auditing of AI and developed either a tool or a framework. The top row shows the author(s), publication year, name of the conference or journal in which the paper was published, and the tool or framework presented. The majority of the selected papers were conference proceedings, followed by journal articles and a few grey literature articles.

The included papers were assessed for suitability vis-à-vis CAAI using the criteria introduced in the following section.

3.2 Assessment Criteria for Continuous Auditing

We derived six assessment criteria from the CA definitions in the literature (Table 2). As presented in Sect. 2.2, continuous AI auditing consists of a continuous system in which processes are repeated regularly and automatically and which has a process for collecting and evaluating data based on predetermined criteria. Studies were given one point for each criterion met, and the total number of points was later used as a baseline when considering suitability for CA. Data collection, automation, and predetermined criteria were the most common criteria met, as they are also

Table 1 Overview of the studied papers

Author	Year	Journal/conference	Output (framework/tool)
Raji et al.	2020	Conference on Fairness, Accountability, and Transparency	Scoping, Mapping, Artifact Collection, Testing, and Reflection framework
AI Ethics Impact Group (AIEIG)	2020	AI Ethics Impact Group	Values, Criteria, Indicators, Observables framework
LaBrie and Steinke	2019	Twenty-Fifth Americas Conference on Information Systems	An Ethical AI Algorithm Audit framework
Bellamy et al.	2019	IBM Journal of Research and Development	AI Fairness 360 toolkit
Cabrera et al.	2019	IEEE Conference on Visual Analytics Science and Technology	FairVis, a visual analytics system
Dawson et al.	2019	Data61 Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australian government	AI: Australia's ethics framework
ECP	2018	ECP Platform for the Information Society	AI impact assessment framework
Epstein et al.	2018	Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence	TuringBox framework
Information Commissioner's Office (ICO)	2020	ICO	AI auditing framework Guidance)
Kim et al.	2019	AAAI/ACM Conference on AI, Ethics, and Society	Multiaccuracy auditing framework
Bird et al.	2020	Microsoft	FairLearn toolkit
Sulaimon et al.	2019	International Conference on Modeling Simulation and Applied Optimization	Control loop framework
Raji et al.	2020	Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society	CelebSET
Black et al.	2020	Conference on Fairness, Accountability, and Transparency	FlipTest
Katell et al.	2020	Conference on Fairness, Accountability, and Transparency	Algorithmic Equity Toolkit
D'Amour et al.	2020	Conference on Fairness, Accountability, and Transparency	An extensible open-source software framework for fairness-focused simulations

Table 1 (continued)

Author	Year	Journal/conference	Output (framework/tool)
Sapiezynski et al.	2019	Companion Proceedings of the 2019 World Wide Web Conference	The Viable-Λ Test for fairness
Brown et al.	2021	Big Data & Society	Auditing framework to guide ethical assessment of an algorithm
Personal Data Protection Commission (PDPC)	2020	Singapore Infocomm Media Development Authority	Model AI governance framework
PricewaterhouseCoopers (PwC)	2019	PwC Responsible AI	Responsible AI toolkit
Saleiro et al.	2018	arXiv	Aequitas bias audit toolkit
Sharma et al.	2019	arXiv	Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of Artificial Intelligence models (CERTIFAI) framework
Smart Dubai	2019	Smart Dubai Office	AI ethics principles and guidelines
World Economic Forum (WEF)	2020	WEF	Facial recognition assessment
Wexler et al. Google	2020	IEEE Transactions on Visualization and Computer Graphics	The What-If Tool
Sutton and Samavi	2018	Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence	A Linked Data-based method of creating tamper-proof privacy audit logs
Javadi et al.	2020	Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society	Indicators for monitoring misuse of "artificial intelligence as a service"
Pasquier et al.	2016	International Conference on Cloud Engineering	Information Flow Audit
Reisman et al.	2018	AI Now algorithmic impact assessment report	A practical framework for public agency accountability
Drakonakis et al.	2020	Conference on Computer and Communications Security	Fully automated black-box auditing framework
Byrnes et al.	2018	AICPA Assurance Services Executive Committee Emerging Assurance Technologies Task Force	Auditor assessment of internal controls
Nandutu et al.	2021	AI & Society	An ethically aligned AI system framework

Table 1 (continued)

Author	Year	Journal/conference	Output (framework/tool)
Panigutti et al.	2021	Information Processing and Management	FairLens, a methodology for discovering and explaining biases
Oala et al.	2021	Journal of Medical Systems	Machine learning for health tool
Lee et al.	2020	Berkeley Technology Law Journal	Risk management framework
Floridi et al.	2022	University of Oxford	capAI procedure
Zicari et al.	2021	IEEE Transactions on Technology and Society	Z-Inspection process
Cobbe et al.	2021	FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency	A Framework for Accountable Algorithmic Systems

Table 2 Criteria derived from continuous auditing definitions

Assessment criterion #	Criterion	Description	Points
AC1	Real time	CA is a (nearly) real-time electronic support system (Eulerich & Kalinichenko, 2018)	+ 1
AC2	Repeat	CA is the property of the function and includes any audit process which is repeated regularly (Marques & Santos, 2017)	+ 1
AC3	Data collection	CA can also be defined as a process of collecting and evaluating data to determine and ensure the efficiency and effectiveness of accounting systems in real time (Marques & Santos, 2017)	+ 1
AC4	Automation	CA continuously and automatically audits clearly defined audit objects based on predetermined criteria (Eulerich & Kalinichenko, 2018)	+ 1
AC5	Predetermined criteria	Continuous assurance may be defined as a process that continually tests transactions and controls based on criteria prescribed by the auditor (Eulerich & Kalinichenko, 2018)	+ 1
AC6	Legal requirement	Auditing is conducted to fulfil legal requirements	+ 1

typical attributes of non-continuous AI auditing. Fulfilment of legal requirements and real-time possibilities followed these most common attributes in prevalence. Audit processes that were repeated regularly were the least used criteria.

Table 3 shows all the studies and how they meet the assessment criteria. "Publication" refers to whether the study was a journal article or conference proceeding (Yes) or a grey literature paper (No). "Type" reflects whether the study develops a tool (T) or a framework (F). By "framework", we mean a conceptual model that presents a set of components and interrelations. In turn, "tool" means a practically applicable tool or set of tools to audit some aspects of AI systems or organizations' use of AI systems. Frameworks were somewhat more common than tools, as 22 studies developed a framework and 13 developed a tool.

Suitability for CA was determined based on the fulfilment of or failure to fulfil the criteria. One point was awarded for the fulfilment of each criterion, after which the points were totalled in the suitability column. Therefore, the more criteria the paper met, the greater its suitability for continuous AI auditing.

4 Findings

The following sections provide an assessment of the auditing tools and frameworks, organized into three clusters: high suitability for CAAI (5–6 points on the criteria introduced in the previous section), medium suitability for CAAI (3–4 points), and low or uncertain suitability for CAAI (0–2 points). Under each cluster, we describe the frameworks and tools currently available in published sources.

4.1 High Suitability for Continuous Auditing of AI (5–6)

The papers that received five or six points based on the assessment were ranked as high-suitability papers for CAAI. This means that these papers either dealt directly with CAAI or satisfied all the CA criteria, making the tools and frameworks they presented suitable for continuous AI auditing, at least provisionally. Seven papers achieved high-suitability status, with six developing a new framework and one developing a tool for CA. The characteristic of the high-suitability papers was that they aimed to define CA or clearly considered its criteria.

The focus of the developed frameworks varied. Lee et al. (2020) and three non-academic papers (Byrnes et al., 2018; ICO, 2020; PDPC, 2020) discussed the evolution of AI and sought to develop guidance for future AI auditing. The rest of the high-suitability papers sought to solve specific problems. For example, D'Amour et al. (2020) provided an open-source software framework for studying the fairness of algorithms, and Pasquier et al. (2016) focused on cloud infrastructure, providing systems to continuously monitor information flows within cloud infrastructure and detect malicious activities and unauthorized changes. Amongst the high-suitability papers, non-academic ones in particular focused on the development of existing AI systems in ways that could be suitable for CAAI. They consider current and future challenges, other AI problems, and how to develop the field in the future.

Table 3 Papers assessed against the continuous auditing criteria

Study	Publication	Type T/F	Real time	Repeat	Data collection	Automation	Predetermined criteria	Legal requirement	Suitability
PDPC	No	F	Yes	Yes	Yes	Yes	Yes	Yes	6
ICO	No	F	Yes	Yes	Yes	Yes	Yes	Yes	6
Floridi et al.	No	T	Yes	Yes	Yes	Yes	Yes	Yes	6
D'Amour et al.	Yes	F	Yes	Yes	Yes	Yes	Yes	No	5
Pasquier et al.	Yes	F	Yes	No	Yes	Yes	Yes	Yes	5
Bird et al.	No	T	Yes	Yes	Yes	Yes	Yes	No	5
Byrnes et al.	No	F	Yes	Yes	Yes	Yes	Yes	No	5
Lee et al.	Yes	F	Yes	No	Yes	Yes	Yes	Yes	5
Dawson et al.	No	F	No	No	Yes	Yes	Yes	Yes	4
Sulaimon et al.	Yes	F	Yes	Yes	Yes	Yes	No	No	4
Brown et al.	Yes	F	No	No	Yes	Yes	Yes	Yes	4
Katell et al.	Yes	T	Yes	No	Yes	Yes	No	Yes	4
Sharma et al.	Yes	T	Yes	Yes	No	Yes	Yes	No	4
Drakonakis et al.	Yes	F	Yes	No	Yes	Yes	Yes	No	4
Zicari et al.	Yes	T	No	No	Yes	No	Yes	Yes	3
Bellamy et al.	Yes	T	Yes	No	Yes	No	Yes	No	3
ECP	No	F	No	No	No	Yes	Yes	Yes	3
Black et al.	Yes	T	Yes	No	Yes	No	No	Yes	3
WEF	No	F	Yes	No	No	Yes	Yes	Yes	3
Raji et al.[a]	Yes	F	No	No	Yes	No	Yes	No	2
AIEIG	No	T	No	No	No	No	Yes	Yes	2
PwC	No	F	No	No	No	No	Yes	Yes	2
Javadi et al.	Yes	F	No	No	No	No	Yes	Yes	2
Nandutu et al.	Yes	F	No	No	Yes	No	No	Yes	2

Table 3 (continued)

Study	Publication	Type T/F	Real time	Repeat	Data collection	Automation	Predetermined criteria	Legal requirement	Suitability
Oala et al.	Yes	F	No	No	No	Yes	Yes	No	2
Cobbe et al.	Yes	F	No	No	No	No	Yes	Yes	2
LaBrie and Steinke	Yes	F	No	No	No	No	Yes	No	1
Cabrera et al.	Yes	T	No	No	No	Yes	No	No	1
Saleiro et al.	Yes	T	No	No	No	No	Yes	No	1
Google	No	T	No	No	No	Yes	No	No	1
Sutton et al.	Yes	T	No	No	No	No	No	Yes	1
Reisman et al.	No	F	No	No	No	Yes	No	No	1
Panigutti et al.	Yes	T	No	No	No	No	Yes	No	1
Epstein et al.	Yes	T	No	No	No	No	No	No	0
Kim et al.	Yes	F	No	No	No	No	No	No	0
Raji et al.[b]	Yes	T	No	No	No	No	No	No	0
Sapiezynski et al.	Yes	F	No	No	No	No	No	No	0
Smart Dubai	No	F	No	No	No	No	No	No	0

[a]Raji et al. (2020a)

[b]Raji et al. (2020b)

Two tools were considered highly suitable for CAAI: FairLearn, developed by Bird et al. (2020), and capAI by Floridi et al. (2022). FairLearn is an open-source toolkit that improves the fairness of AI systems. It has an interactive visualization dashboard and unfairness mitigation algorithms that manage trade-offs between fairness and model performance. The goal of FairLearn is to mitigate fairness-related harm. Fully achieving guaranteed fairness is challenging, as societal and technical systems are highly complex. FairLearn recognizes a wide range of fairness-related harms and ways to improve fairness and detect unfair activities. For example, an AI system can unfairly allocate opportunities, resources, or information or fail to provide all people with the same quality of service. In addition, it can reinforce existing stereotypes, denigrate people, or overrepresent or underrepresent groups of people. The aim of FairLearn is to address the gap in software, thus tackling these fairness issues continuously and focusing in particular on negative impacts.

The main purpose of capAI (Floridi et al., 2022), in turn, is to serve as a governance tool. It aims to ensure conformity with the EU's Artificial Intelligence Act by demonstrating that AI systems are developed and operated in a trustworthy manner. CapAI views AI systems across the entire AI life cycle from design to retirement. It defines and reviews current practices and enables technology providers and users to develop ethical assessments at each stage of the AI life cycle. The procedure consists of an internal review protocol, an external scorecard, and a summary data sheet. These allow organizations to conduct conformity assessments and the technical documentation required by the AIA. They produce a high-level summary of the AI system's purpose, functionality, and performance and summarize relevant information about the AI system's purpose, values, data, and governance (Floridi et al., 2022).

Overall, the high-suitability papers dealt with developing automated CA systems that met each assessment criterion. Therefore, even if the paper did not explicitly develop a framework or tool for CAAI, it was considered suitable for this purpose. An automated and continuous system, which is repeated regularly, was an essential aspect of most frameworks and tools. ICO (2020), PDPC (2020), and Byrnes et al. (2018), all of which touched on the future of AI, noticed themes arising in relation to CA. Interestingly, only one tool was developed in high-suitability papers. This could indicate that the discussion is centred more on the general definition and direction of CAAI than on the development of new tools.

4.2 Medium Suitability for Continuous Auditing of AI (3–4)

Ten papers received three or four points from our assessment: Six were frameworks, and four were tools. The most typical characteristics of the medium-suitability papers were real time, data collection, automation, and predetermined criteria. "Repeat" was clearly the least common criterion fulfilled, followed by "legal requirements". This indicates that medium-suitability papers may be suitable for continuous AI auditing, but in principle, they were not designed for CA. However, as seven of the 10 medium-suitability papers matched the real-time criteria, it can be stated that the

difference between medium- and high-suitability papers is small in practice and that medium-suitability papers are also relevant to continuous AI auditing.

The tools and frameworks with medium suitability are divided into ethics-based frameworks and technical approaches to specific problems. On the ethics-based side, Brown et al. (2021) presented an auditing framework to guide the ethical assessment of an algorithm. Regarding the non-academic papers in the medium-suitability category, there were many similarities with the non-academic papers in the high-suitability category. In *Artificial Intelligence: Australia's Ethics Framework*, Dawson et al. (2019) covered civilian applications of AI with the goal of developing best practice guidelines. Similarly, the Dutch information society platform ECP (2018) wrote an AI impact assessment framework to build guidelines for the rules of conduct of autonomous systems, and the WEF (2020) wrote a policy framework addressing responsible limits regarding facial recognition. All these frameworks cover ethical aspects of the development of AI, taking into account the characteristics of AI systems, but real-time capabilities and the repeated nature of procedures are given less attention.

Amongst the technical approaches were, for instance auditing frameworks focusing on black-box auditing and bias. The automation of activities was the focus of the fully automated black-box auditing framework by Drakonakis et al. (2020). The framework aims to detect authentication and authorization flaws when handling cookies that stem from the incorrect, incomplete, or non-existent deployment of appropriate security mechanisms. Sulaimon et al. (2019) and Thangavel et al. (2020) focused on security, bias, and data issues. Sulaimon et al. (2019) proposed a control loop, which is an adaptation of the Monitor, Analyse, Plan, Execute, and Knowledge control loop for autonomous systems. Their goal is to ensure fairness in the decision-making processes of automated systems by adapting the existing bias detection mechanism. Thangavel et al. (2020) also aimed to develop existing systems to increase and maintain cloud users' trust in cloud service providers. They proposed a novel integrity verification framework that performs block-, file-, and replica-level auditing to verify data integrity and ensure data availability in the cloud.

Ethical considerations played an essential role in the tools presented in the medium-suitability papers. AI Fairness 360 by Bellamy et al. (2019) and FlipTest by Black et al. (2020) focused on fairness issues in AI systems. Their main objective is to help facilitate fairness algorithms for users to progress as seamlessly as possible from raw data to a fair model. Greater fairness, accountability, and transparency in algorithmic systems were also the objectives of the Algorithmic Equity Toolkit by Katell et al. (2020) and the Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of Artificial Intelligence models (CERTIFAI; Sharma et al., 2019). Zicari et al. (2021) assessed AI trustworthiness by developing the Z-Inspection process that assesses and seeks to resolve ethical issues and tensions in AI usage domains.

In summary, medium-suitability papers offer important guidelines and tools for continuous AI auditing. CA was not the core focus of the papers, but similarities and applicability to continuous AI auditing were seen. In particular, continuous and real-time opportunities were a point of interest in the medium-suitability papers. However, systems which operate repeatedly and automatically did not

stand out as strongly as they did in papers that received five or six suitability assessment points. Additionally, the medium-suitability papers did not recognize or define the concept of CA as clearly as the high-suitability papers did.

4.3 Low or Uncertain Suitability for Continuous Auditing of AI (0–2)

Papers that received 0–2 points from the assessment were ranked as low- or uncertain-suitability papers. Eighteen papers were considered to have low or uncertain suitability; this was clearly the largest category amongst the studied papers. The low suitability score means that the CA criteria were neither mentioned nor specified in these papers; in particular, real-time and repeat criteria were not found. The most common criteria fulfilled in this category were "predetermined criteria" and "legal requirement", followed by "data collection". Owing to their low suitability for CA, we do not discuss these papers in detail. However, it should be noted that some frameworks and tools could nevertheless be adapted to suit CA. For example, formulating guidelines for ethical AI auditing, bringing principles into practice, and designing tools for specific issues might bring significant insight into the continuous AI auditing discussion, even though the framework or tool itself is not intended for CA.

5 Discussion and Conclusion

We draw out the central implications of our conceptualization and assessment of CAAI frameworks and tools in the following sections. First, we lay out the state of the art in CAAI. Then, we point to lessons from existing CA frameworks from fields other than AI. Finally, we discuss the central problem of automation and human oversight, and we conclude the paper with limitations and future research directions.

5.1 The State of the Art in Continuous Auditing of AI

CAAI is an emerging field, and we are only beginning to draw its contours. Whilst we were able to find literature on auditing of AI and CA, none explicitly connected these two topics as the core focus of the paper. At the same time, based on our overview, there is significant potential for continuous approaches to the auditing of AI. In the following paragraphs, we present provisional rather than definite conclusions because the area is moving quickly and many frameworks and tools may have untapped potential.

To sum up the findings in the previous section, no clear pattern is emerging from the high-suitability audit tools and frameworks. They are highly heterogeneous software frameworks, risk management frameworks, and other auditing tools. Considering the criteria for judging CA, it seems that the automated, real-time, and repeated nature of auditing are essential criteria for the continuous nature of AI auditing. The remaining criteria (data collection, predetermined criteria, and legal requirements)

condition the specific type of auditing and are part of traditional non-continuous auditing.

Given the early stage of the conceptualization of CAAI, it is useful to consider the basic distinctions between potential CAAI tools. One clear difference is between sector-specific tools (e.g. healthcare) and cross-sectoral tools. There is a trade-off between sector-specific tools that can focus on sectorally relevant issues (e.g. privacy issues in healthcare) and general tools that are more abstract and may either leave out sectorally important AI governance issues or include irrelevant issues. Another crucial axis is the desired level of automation in the overall auditing process. With a comparatively low level of automation, CAAI can assist auditors and provide additional information on the fairness of algorithms, for instance. If the desired level of automation is high, the auditing process can be automated to a large extent. Then, the human auditor has a more limited role akin to the "human-on-the-loop" model, whereby automated systems can make decisions, but a human oversees them and intervenes in the case of incorrect decisions (Benjamins, 2021).

Our study made the distinction between frameworks and tools, which may be difficult to discern in practice. Going forward, we hypothesize that both general frameworks and specialized tools are needed in CAAI. Practical tools are likely to be most valuable when used as part of a more general auditing framework that contextualizes the tools and the information they provide. CAAI could thus be seen as a nested system with an overarching framework and a set of specific tools under this framework.

Considering recent developments in AI regulation, one strong candidate for an overarching general framework is the proposed EU AI Act (European Commission, 2021). The AI Act proposal includes provisions for the mandatory post-market monitoring of high-risk AI systems, which requires AI system providers to draft post-market monitoring plans to document the performance of high-risk AI systems throughout their life cycles (Mökander et al., 2022). At present, however, the AI Act leaves the practical implementation of post-market monitoring largely open (Mökander et al., 2022). This is where CAAI frameworks and tools could contribute by concretizing the generic AI Act and providing practical tools for AI developer organizations. The EU AI Act may thus increase the demand for CAAI tools. At the same time, CAAI tools can also offer help for ethics-based AI auditing in areas not covered by the EU AI Act, including lower risk systems and broader ethical and impact issues. In other words, CAAI tools could supplement legally binding requirements and support corporate social responsibility and business ethics.

To take the CAAI field forward, it is useful to look at the possible types of CAAI differentiated by their maturity. Here, we draw on Gregor and Hevner's (2013) distinction between two axes—solution maturity (low/high) and application domain maturity (low/high)—which were developed to enable the understanding of different types of design science contributions. These axes yield four possible types of CAAI frameworks and tools: improvement (new solutions for known problems), invention (new solutions for new problems), routine design (known solutions for known problems), and exaptation (extending known solutions to new problems). Table 4 lays out these different types.

Based on our assessment of frameworks and tools, the CAAI field is not yet in the stage of routine design because CAAI solutions are still emerging. According to

Table 4 Solution maturity and application domain maturity in continuous auditing of artificial intelligence (based on Gregor & Hevner, 2013)

	High-application domain maturity	**Low-application domain maturity**
Low-solution maturity	**Improvement:** develop new CAAI solutions for known problems	**Invention:** invent new CAAI solutions for new problems
High-solution maturity	**Routine design:** apply known CAAI solutions to known problems	**Exaptation:** extend known CAAI solutions to new problems (e.g. adopt solutions from other fields)

a review of the design science literature, invention, the generation of new solutions for new problems, is rare in practice (Gregor & Hevner, 2013). However, there is significant scope for improvement and exaptation in the CAAI field. The present study has focused largely on improvement—that is, the development of new CAAI solutions for known problems. This means that the problems—regarding AI ethics and trustworthiness, for example—are known, but new tools are needed to improve the maturity of the solutions. The final category, exaptation, means that existing solutions would be extended to new problem areas. In this case, it means applying CA frameworks from other fields to auditing of AI. This is a promising direction that complements our study of existing AI auditing solutions by looking at CA solutions and asking what could be learned regarding the auditing of AI. We turn to this question in the following section.

5.2 Drawing Lessons for AI Auditing From the Existing Continuous Auditing Frameworks

This study assessed frameworks and tools intended for auditing of AI in light of their suitability for CA. A logical next step is to approach the issue from the opposite direction and draw lessons from existing frameworks developed for the CA of entities other than AI systems. How can aspects of CA frameworks be adapted to audit AI systems?

The CA literature presents numerous CA frameworks intended for financial and IT auditing. For example, Yoon et al. (2021) and Shiue et al. (2021) developed frameworks for CA systems. Yoon et al. (2021) presented a CA system with alarms for unusual transactions and exceptions on three levels. Shiue et al.'s (2021) work explores key criteria for implementing CA systems based on two approaches: an embedded audit module and a monitoring and control layer. Going further, Majdalawieh et al. (2012) designed a full-power CA model which supports business process-centric auditing and business monitoring whilst enabling the fulfilment of compliance requirements within internal and external policies and regulations. Their model has three objectives: build a CA model on the principle of continuous monitoring and with predefined components, facilitate the integration of CA and business information processing within an enterprise using different building blocks, and give practitioners insight into the state of the adoption of CA in the enterprise and how it will enhance their audit effectiveness and audit efficiency. Tronto and Killingsworth (2021) also focused on developing a continuous monitoring tool for collaboration between internal auditing and business operations. Kiesow et al. (2014) recognized the problems with the implementation of CA and noted that traditional audit tools neglect the potential of Big Data analytics. Therefore, they strived to develop a computer-assisted audit solution. Wang et al. (2020) proposed a continuous compliance awareness framework to audit an organization's purchases in an automatic and timely manner. Eulerich et al. (2022) developed a three-step evaluation framework to facilitate robotic process automation and assist auditors in deciding what activities should be automated.

Common to these existing CA solutions is that they are organized around business and accounting processes, such as purchase orders and invoices. In contrast, auditing of AI focuses on auditing an AI system's consistency with relevant norms

and standards. Owing to this difference in focus, the existing CA frameworks cannot be directly adopted as CAAI frameworks; instead, they need to be adapted to serve as CAAI solutions. The further development of their adaptation is beyond the scope of this paper, but we offer three points that are relevant to this future work:

- The scope of CAAI needs to be specified, particularly whether CAAI should focus on a specific algorithmic system or if it extends more broadly to auditing organizations' use of AI systems in the future.
- The human–machine division of labour needs to be considered to define which aspects of AI auditing should be automated and which should not.
- Emerging CAAI systems must be considered in light of the emerging actor landscape and institutions of AI governance, such as a possible AI regulatory body in the European Union (Stix, forthcoming).

5.3 Automation and Human Oversight

Whilst the automation of auditing promises efficiency, there is a risk of introducing a second-order problem. If opaque and unpredictable automated systems are the original problem, can automated auditing also become opaque and unpredictable? The assurance of AI systems could thus lead to a kind of infinite regress: the systems that audit AI systems need to be audited, the systems that audit the auditing systems need to be audited, and so on. As an organizational response to this problem, the established "three lines of defence" model, which includes operational management, risk management functions, and internal audits, could be adapted to manage AI risks (Institute of Internal Auditors, 2020; cf. Financial Services Agency of Japan, 2021). Addressing a similar problem, Metcalf et al. (2021) wrote about "the assessor's regress" in the context of impact assessments, whereby the completeness of an assessment relies on a never-ending chain of justification. Their answer to this dilemma is that a forum and a legitimate accountability relationship must exist to close the regress. In the CAAI context, mechanisms for creating trust in the auditing system are needed. However, exploring more details about such mechanisms and the connections to the three lines of defence model is beyond the scope of this paper.

On a broader societal level, CA raises a challenge regarding the widely accepted notions of the human oversight of AI systems (Floridi et al., 2018; Shneiderman, 2020). Ensuring human oversight, human-centricity, and agency over opaque AI systems is one of the central principles of AI ethics (Dignum, 2020; High-Level Expert Group on Artificial Intelligence, 2019). Against this background, CAAI can be seen to diminish human control and understandability in the auditing process because part of auditing work is transferred to machines.

However, there is another possible reading of CAAI from a human-centric AI perspective. It can be argued that outsourcing part of the mechanical auditing work to machines frees human auditors to focus on higher level auditing and oversight tasks. If CAAI is designed in a human-centric manner, it can augment rather than diminish human capabilities. Transferring oversight tasks from humans to machines

can paradoxically increase human oversight of AI systems, but this requires an appropriate CAAI design.

The general conclusion from this discussion is that CAAI systems should be kept relatively simple and transparent to avoid adding layers of opaqueness and complexity to already complex systems. In this case, the assurance of CA is more straightforward than the assurance of complex algorithmic systems. Another potential solution to the second-order assurance problem is the standardization and issuing of certifications for CA products to create trust in CA. At least initially, we can assume that the assurance of CA processes is more straightforward than the assurance of complex AI systems and the assessment of their societal impacts.

5.4 Limitations and Future Research Directions

As a foray into a novel topic, this study has some limitations. It is still too early to conduct a systematic literature review specifically on CAAI; hence, our assessment's coverage of relevant publications, frameworks, and tools may be incomplete. However, this is a challenge with any discussion on a fast-moving topic, such as AI auditing, in which technologies and legislation continuously co-evolve. Moreover, our study does not cover the technical aspects and processual details of CAAI. In other words, we do not delve into the complexities of gaining visibility into black-box systems. Further technical and organizational research is likely needed for CAAI to be practically feasible.

Owing to its exploratory nature, this study suggests significant areas for future research. As CAAI frameworks and tools mature, a systematic literature review will become a helpful tool for gaining a bird's-eye view of the developing field. In addition, studies could drill down into sectoral requirements and actor dynamics in particular industries, such as healthcare, public administration, and finance. The interplay between sectoral legislation, generic AI legislation, and ethical and stakeholder requirements provides rich avenues for case studies; comparative studies; and, eventually, quantitative studies on a larger scale.

Funding Open Access funding provided by University of Turku (UTU) including Turku University Central Hospital. The research was conducted in the Artificial Intelligence Governance and Auditing (AIGA) project funded by Business Finland.

Data Availability The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest The authors declare no competing interests.

References

AI Ethics Impact Group. (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

American Institute of Certified Public Accountants. (1999). *Continuous auditing research report*. American Institute of Certified Public Accountants.

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilović, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development, 63*(4/5), 4:1–4:15. https://doi.org/10.1147/JRD.2019.2942287

Benjamins, R. (2021). A choices framework for the responsible use of AI. *AI and Ethics, 1*(1), 49–53. https://doi.org/10.1007/s43681-020-00012-5

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., Walker, K., & Design, A. (2020). *Fairlearn: A toolkit for assessing and improving fairness in AI*. 7.

Black, E., Yeom, S., & Fredrikson, M. (2020). FlipTest: Fairness testing via optimal transport. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 111–121. https://doi.org/10.1145/3351095.3372845

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

Byrnes, P. E., Al-Awadhi, A., Gullvist, B., Brown-Liburd, H., Teeter, R., Warren, J. D., & Vasarhelyi, M. (2018). Evolution of auditing: From the traditional approach to the future audit. In D. Y. Chan, V. Chiu, & M. A. Vasarhelyi (Eds.), Continuous Auditing (pp. 285–297). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78743-413-420181014

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*. https://doi.org/10.48550/ARXIV.1904.05419

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 598–609. https://doi.org/10.1145/3442188.3445921

Coderre, D. (2005). Continuous auditing: Implications for assurance, monitoring, and risk assessment. *Global technology audit guide*. The Institute of Internal Auditors.

D'Amour, A., Srinivasan, H., Atwood, J., Baljekar, P., Sculley, D., & Halpern, Y. (2020). Fairness is not static: Deeper understanding of long term fairness via simulation studies. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 525–534. https://doi.org/10.1145/3351095.3372878

Dawson, D., & Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., Scowcroft, J., & Hajkowicz, S. (2019). Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia. Retrieved February 11, 2021, from https://www.csiro.au/en/research/technology-space/ai/AIEthics-Framework

Dignum, V. (2020). Responsibility and artificial intelligence. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 213–231). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.12

Doneda, D., & Almeida, V. A. F. (2016). What is algorithm governance? *IEEE Internet Computing, 20*(4), 60–63. https://doi.org/10.1109/MIC.2016.79

Drakonakis, K., Ioannidis, S., & Polakis, J. (2020). The Cookie Hunter: Automated black-box auditing for web authentication and authorization flaws. *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 1953–1970. https://doi.org/10.1145/3372297.3417869

ECP. (2018). Artificial Intelligence Impact Assessment (English version). Retrieved February 20, 2021, from https://ecp.nl/publicatie/artificial-intelligence-impactassessment-english-version/

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 systems. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 5826–5828. https://doi.org/10.24963/ijcai.2018/851

Eulerich, M., & Kalinichenko, A. (2018). The current state and future directions of continuous auditing research: An analysis of the existing literature. *Journal of Information Systems, 32*(3), 31–51. https://doi.org/10.2308/isys-51813

Eulerich, M., Pawlowski, J., Waddoups, N. J., & Wood, D. A. (2022). A framework for using robotic process automation for audit tasks. *Contemporary Accounting Research*, *39*(1), 691–720. https://doi.org/10.1111/1911-3846.12723

European Commission. (2021). *Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts com/2021/206 final.* Retrieved August 1, 2022, from https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence

Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., Eling, M., Goodloe, A., Gupta, J., Hart, C., Jirotka, M., Johnson, H., LaPointe, C., Llorens, A. J., Mackworth, A. K., Maple, C., Pálsson, S. E., Pasquale, F., Winfield, A., & Yeong, Z. K. (2021). Governing AI safety through independent audits. *Nature Machine Intelligence, 3*(7), 566–571. https://doi.org/10.1038/s42256-021-00370-7

Financial Services Agency of Japan. (2021). *Principles for model risk management*. https://www.fsa.go.jp/common/law/ginkou/pdf_03.pdf

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mökander, J., & Wen, Y. (2022). *CapAI—A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act* (SSRN Scholarly Paper ID 4064091). *Social Science Research Network*. https://doi.org/10.2139/ssrn.4064091

Galdon Clavell, G., Martín Zamorano, M., Castillo, C., Smith, O., & Matic, A. (2020, February). Auditing algorithms: On lessons learned and the risks of data minimization. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 265-271. https://doi.org/10.1145/3375627.3375852

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Groomer, S. M., & Murthy, U. S. (1989). Continuous auditing of database applications: An embedded audit module approach. *Journal of Information Systems, 3*(2), 53.

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. Retrieved September 10, 2020, from https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

Information Commissioner's Office. (2020). Guidance on the AI auditing framework: Draft guidance for consultation. Retrieved February 11, 2021, from https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

Institute of Internal Auditors. (2020). *The IIA's three lines model: An update of the three lines of defense*. Retrieved August 1, 2022, from https://www.theiia.org/globalassets/site/about-us/advocacy/three-lines-model-updated.pdf

Institute of Internal Auditors. (2022). *About internal audit*. Retrieved August 22, 2022, from https://www.theiia.org/en/about-us/about-internal-audit/

Javadi, S. A., Cloete, R., Cobbe, J., Lee, M. S. A., & Singh, J. (2020). Monitoring Misuse for Accountable'Artificial Intelligence as a Service'. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 300-306. https://doi.org/10.1145/3375627.3375873

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Bintz, C., Raz, D., & Krafft, P. M. (2020). Toward situated interventions for algorithmic equity: Lessons from the field. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 45–55. https://doi.org/10.1145/3351095.3372874

Kiesow, A., Zarvic, N., & Thomas, O. (2014). Continuous auditing in big data computing environments: Towards an integrated audit approach by using CAATTs. GI-Jahrestagung.

Kim, M. P., Ghorbani, A., & Zou, J. (2019). Multiaccuracy: Black-box post-processing for fairness in classification. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 247–254. https://doi.org/10.1145/3306618.3314287

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting*, *14*(1), 115–122. Scopus. https://doi.org/10.2308/jeta-51730

Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., & Lomas, E. (2021). *Towards algorithm auditing: A survey on managing legal, ethical and technological risks of AI, ML and associated algorithms* (SSRN Scholarly Paper ID 3778998). *Social Science Research Network*. https://doi.org/10.2139/ssrn.3778998

Laato, S., Birkstedt, T., Mäntymäki, M., Minkkinen, M., & Mikkonen, T. (2022a). AI governance in the system development life cycle: Insights on responsible machine learning engineering. *Proceedings of the 1st Conference on AI Engineering—Software Engineering for AI*.

Laato, S., Mäntymäki, M., Minkkinen, M., Birkstedt, T., Islam, A. K. M. N., & Dennehy, D. (2022b). Integrating machine learning with software development lifecycles: Insights from experts. *ECIS 2022b Proceedings*. ECIS, Timişoara, Romania.

LaBrie, R., & Steinke, G. (2019). Towards a Framework for Ethical Audits of AI Algorithms. AMCIS 2019 Proceedings. https://aisel.aisnet.org/amcis2019/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/24

Lee, M. S. Ah., Floridi, L., & Denev, A. (2020). Innovating with confidence: Embedding AI governance and fairness in a financial services risk management framework. In L. Floridi (Ed.), *Ethics, governance, and policies in artificial intelligence* (Vol. 144, pp. 353–371). Springer International Publishing. https://doi.org/10.1007/978-3-030-81907-1_20

Majdalawieh, M., Sahraoui, S., & Barkhi, R. (2012). Intra/inter process continuous auditing (IIPCA), integrating CA within an enterprise system environment. *Business Process Management Journal, 18*(2), 304–327. https://doi.org/10.1108/14637151211225216

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022a). Defining organizational AI governance. *AI and Ethics*. https://doi.org/10.1007/s43681-022-00143-x

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022b). *Putting AI ethics into practice: The hourglass model of organizational AI governance* (arXiv:2206.00335). arXiv. https://doi.org/10.48550/arXiv.2206.00335

Marques, R. P., & Santos, C. (2017). Research on continuous auditing: A bibliometric analysis. *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–4. https://doi.org/10.23919/CISTI.2017.7976048

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

Metcalf, J., Moss, E., Watkins, E. A., Singh, R., & Elish, M. C. (2021). Algorithmic impact assessments and accountability: The co-construction of impacts. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 735–746. https://doi.org/10.1145/3442188.3445935

Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2022a). *What about investors?* AI & SOCIETY. https://doi.org/10.1007/s00146-022-01415-0

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022b). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10269-2

Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2022). Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds and Machines, 32*, 241–268. https://doi.org/10.1007/s11023-021-09577-4

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics, 27*(4), 44. https://doi.org/10.1007/s11948-021-00319-4

Nandutu, I., Atemkeng, M., & Okouma, P. (2021). Integrating AI ethics in wildlife conservation AI systems in South Africa: A review, challenges, and future research agenda. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01285-y

Oala, L., Murchison, A. G., Balachandran, P., Choudhary, S., Fehr, J., Leite, A. W., Goldschmidt, P. G., Johner, C., Schörverth, E. D. M., Nakasi, R., Meyer, M., Cabitza, F., Baird, P., Prabhu, C., Weicken,

E., Liu, X., Wenzel, M., Vogler, S., Akogo, D., & Wiegand, T. (2021). Machine learning for health: Algorithm auditing & quality control. *Journal of Medical Systems, 45*(12), 105. https://doi.org/10.1007/s10916-021-01783-y

Panigutti, C., Perotti, A., Panisson, A., Bajardi, P., & Pedreschi, D. (2021). FairLens: Auditing black-box clinical decision support systems. *Information Processing & Management, 58*(5), 102657. https://doi.org/10.1016/j.ipm.2021.102657

Pasquale, F. (2015). *The black box society: The secret al.gorithms that control money and information*. Harvard University Press.

Pasquier, T. F. J.-M., Singh, J., Bacon, J., & Eyers, D. (2016). Information flow audit for PaaS clouds. *2016 IEEE International Conference on Cloud Engineering (IC2E)*, 42–51. https://doi.org/10.1109/IC2E.2016.19

PDPC. (2020). PDPC Model AI Governance Framework, Second Edition. Retrieved February 11, 2021, from https://iapp.org/resources/article/pdpc-model-ai-governance-framework-second-edition/

PwC. (2019). Responsible AI Toolkit. Retrieved August 1, 2022, from https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-isresponsible-ai.html

Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., & Denton, E. (2020). Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 145-151. https://doi.org/10.1145/3375627.3375820

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020b). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020b Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic impact assessments: A practical framework for public agency accountability*. AI Now. Retrieved August 22, 2022, from http://www.tandfonline.com/doi/abs/10.1080/07349165.1995.9726076

Rezende, I. N. (2020). Facial recognition in police hands: Assessing the 'Clearview case' from a European perspective. *New Journal of European Criminal Law, 11*(3), 375–389.

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Saleiro, P., Kuester, B., Hinkson, L., London, J., Stevens, A., Anisfeld, A., Rodolfa, K. T., & Ghani, R. (2018). *Aequitas: A bias and fairness audit toolkit*. https://doi.org/10.48550/ARXIV.1811.05577

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: Converting critical concerns into productive inquiry: A preconference at the 64th Annual Meeting of the International Communication Association*.

Sapiezynski, P., Zeng, W., E Robertson, R., Mislove, A., & Wilson, C. (2019). Quantifying the impact of user attention on fair group representation in ranked lists. *Companion proceedings of the 2019 World Wide Web Conference*, 553–562. https://doi.org/10.1145/3308560.3317595

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2022). Artificial intelligence governance for businesses. *Information Systems Management*. https://doi.org/10.1080/10580530.2022.2085825

Seppälä, A., Birkstedt, T., & Mäntymäki, M. (2021). From ethical AI principles to governed AI. *Proceedings of the 42nd International Conference on Information Systems (ICIS2021)*. International Conference on Information Systems (ICIS), Austin, Texas. Retrieved March 3, 2022, from https://aisel.aisnet.org/icis2021/ai_business/ai_business/10/

Shah, H. (2018). Algorithmic accountability. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences, 376*(2128), 20170362. https://doi.org/10.1098/rsta.2017.0362

Sharma, S., Henderson, J., & Ghosh, J. (2019). *CERTIFAI: Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of Artificial Intelligence models*. https://doi.org/10.48550/ARXIV.1905.07857

Shiue, W., Liu, J. Y., & Li, Z. Y. (2021). Strategic multiple criteria group decision-making model for continuous auditing system. *Journal of Multi-Criteria Decision Analysis, 28*(5–6), 269–282. https://doi.org/10.1002/mcda.1758

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems, 10*(4), 26. https://doi.org/10.1145/3419764

Smart Dubai. (2019). AI ethics principles and guidelines. Retrieved August 1, 2022, from https://www.digitaldubai.ae/docs/default-source/ai-principlesresources/ai-ethics.pdf

Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. Social Studies of Science, 48(1), 25–56. https://doi.org/10.1177/0306312717741687

Stix, C. (forthcoming). The ghost of AI governance past, present and future: AI governance in the European Union. In J. Bullock & V. Hudson (Eds.), *Oxford University Press handbook on AI governance*. Oxford University Press.

Sulaimon, I. A., Ghoneim, A., & Alrashoud, M. (2019). A new reinforcement learning-based framework for unbiased autonomous software systems. *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*, 1–6. https://doi.org/10.1109/ICMSAO.2019.8880288

Sutton, A., & Samavi, R. (2018). Tamper-proof privacy auditing for artificial intelligence systems. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 5374–5378. https://doi.org/10.24963/ijcai.2018/756

Tewari, G. (2022). Council post: The future of AI: 5 things to expect in the next 10 years. *Forbes*. Retrieved August 11, 2022, from https://www.forbes.com/sites/forbesbusinesscouncil/2022/05/05/the-future-of-ai-5-things-to-expect-in-the-next-10-years/

Thangavel, M., & Varalakshmi, P. (2020). Enabling Ternary Hash Tree Based Integrity Verification for Secure Cloud Data Storage. *IEEE Transactions on Knowledge and Data Engineering*, *32*(12), 2351–2362. https://doi.org/10.1109/TKDE.2019.2922357

Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2021). Responsible AI for digital health: A synthesis and a research agenda. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10146-4

Tronto, S., & Killingsworth, B. L. (2021). How internal audit can champion continuous monitoring in a business operation via visual reporting and overcome barriers to success. *The International Journal of Digital Accounting Research*, *21*(27), 23-59. https://doi.org/10.4192/1577-8517-v21_2

Vasarhelyi, M. A., & Halper, F. (1991). The continuous audit of online systems. *Auditing: A Journal of Practice & Theory*, *10*(1).

Wang, K., Zipperle, M., Becherer, M., Gottwalt, F., & Zhang, Y. (2020). An AI-based automated continuous compliance awareness framework (CoCAF) for procurement auditing. *Big Data and Cognitive Computing, 4*(3), 23. https://doi.org/10.3390/bdcc4030023

WEF (World Economic Forum). (2020). A Framework for Responsible Limits on Facial Recognition Use Case: Flow Management. Retrieved February 20,2021, from http://www3.weforum.org/docs/WEF_Framework_for_action_Facial_recognition_2020.pdf

Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viégas, F., & Wilson, J. (2020). The What-If Tool: Interactive Probing of Machine Learning Models. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 56–65. https://doi.org/10.1109/TVCG.2019.2934619

Yeung, K., Howes, A., & Pogrebna, G. (2020). AI governance by human rights-centered design, deliberation, and oversight: An end to ethics washing. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 75–106). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.5

Yoon, K., Liu, Y., Chiu, T., & Vasarhelyi, M. A. (2021). Design and evaluation of an advanced continuous data level auditing system: A three-layer structure. *International Journal of Accounting Information Systems*, *42*, 100524. https://doi.org/10.1016/j.accinf.2021.100524

Zicari, R. V., Brodersen, J., Brusseau, J., Dudder, B., Eichhorn, T., Ivanov, T., Kararigas, G., Kringen, P., McCullough, M., Moslein, F., Mushtaq, N., Roig, G., Sturtz, N., Tolle, K., Tithi, J. J., van Halem, I., & Westerlund, M. (2021). Z-Inspection: A process to assess trustworthy AI. *IEEE Transactions on Technology and Society, 2*(2), 83–97. https://doi.org/10.1109/TTS.2021.3066209