Bridging Data, Intelligence, and Trust the Future of Computational Systems and Ethical AI

Rohit Yallavula[1*], Siva Karthik Parimi[2]

[1] Independent Researcher, University of Texas, Dallas, TX, UNITED STATES
[2] Senior Software Engineer, PayPal, Austin, TX, UNITED STATES

*Corresponding Author Email: Rohit.yallavula07@gmail.com

ABSTRACT

The issue of ethical governance for robotics and artificial intelligence (AI) systems is examined in this paper. As a framework to guide ethical governance in robotics and AI, we present a roadmap that connects several components, such as ethics, standards, regulation, responsible research and innovation, and public engagement. We conclude by suggesting five pillars of sound ethical governance, arguing that ethical governance is crucial to fostering public confidence in robotics and AI. The theme issue "Governing artificial intelligence: ethical, legal, and technical opportunities and challenges" includes this article. In order to establish and preserve public trust and guarantee that robotics and artificial intelligence (AI) systems are created for the benefit of the general public, this paper aims to make the case for a more inclusive, open, and flexible system of governance. It is crucial to increase public confidence in intelligent autonomous systems (IAS). The benefits of IAS to society and the economy will not materialise without that trust. We will lay out a roadmap in this paper. A roadmap serves two purposes: first, it links and charts the various components that make up IAS ethics; second, it gives us a framework to direct ethical governance.

Keywords: Data Governance; Ethical AI; Privacy-Preserving Computation; Federated Learning; Responsible Data Management; Trust Frameworks.

Introduction

In order to fulfil roles in the workplace, home, leisure, healthcare, social care, and education, the private sector has invested heavily over the past ten years in the development of autonomous robots and artificial intelligence (AI) that can interact with humans. These advancements could have enormous positive effects on society. They can cut expenses, save time, and require less human labour to complete tasks. Additionally, they can enhance well-being by offering trustworthy care assistance to the ageing population, standardising service interactions, providing companionship and affective aids to various user groups, and relieving people of both hazardous and mundane tasks.

The general public's perception of these new intelligent technologies is favourable [1,2]. Concerns have been expressed, nevertheless, about the careless application and possibly detrimental effects of IAS. These issues are commonly brought up in public discourse, which frequently frames the future pervasiveness of robots as inevitable and creates bleak scenarios in which human autonomy, security, and authority are usurped. However, there are legitimate concerns about the impact on jobs and mass unemployment, which may be

well-founded [3]. We know that there is no "formula" for establishing trust, but we also know from experience that if technology is safe, well-regulated, and beneficial, people will generally trust it.

A variety of strategies, from those at the level of individual systems and application domains [4] to those at the institutional level [5,6], will be needed to establish such trust in robotics and AI. This essay makes the case that ethical governance is a crucial (though not sufficient) component in fostering trust in IAS. We characterise ethical governance as a collection of policies, practices, values, and cultures intended to guarantee the highest standards of conduct. Therefore, ethical governance is more than just good (i.e., effective) governance because it instills ethical behaviour in both individual designers and the organisations where they are employed. A key component of responsible research and innovation (RI) is normative ethical governance, which "entails an approach, rather than a mechanism, so it seeks to deal with ethical issues as or before they arise in a principled manner rather than waiting until a problem surfaces and dealing with it in an ad hoc way" [7].

New and flexible governance procedures are required in light of the accelerating rate of innovation [8]. The quick speed of revolutionary technological innovation is "reshaping industries, blurring geographical boundaries and challenging existing regulatory frameworks," according to a recent World Economic Forum (WEF) white paper [9]. The report advocates for a more inclusive and flexible form of governance, noting that businesses and innovators, in addition to policymakers, feel obligated to engage with policies to address the societal consequences of their innovations. In order to rethink policy-making for the fourth industrial revolution, the World Economic Forum (WEF), an international organisation for public-private cooperation, is initiating a global agile governance initiative [10]. "Adaptive, human-centered, inclusive, and sustainable policy-making, which acknowledges that policy development is no longer limited to governments but rather is an increasingly multi-stakeholder effort," is how the WEF defines agile governance [11]. The key to making ethical governance both flexible and realistic is the involvement of non-governmental stakeholders, such as individual researchers, research institutions and funders, professional associations, industry, and civil society. In actuality, this entails integrating various types of knowledge—including citizen knowledge—to guide the objectives and paths of innovation.

The ethical governance of both software AIs, such as personal digital assistants or medical diagnosis AIs, and physical robots, such as driverless cars or personal assistant robots (for care or the workplace), is the focus of this paper. We refer to all of these as "intelligent autonomous systems" (IAS) since they are intelligent agents with varying degrees of autonomy. New ethical guidelines for robots and artificial intelligence (AI) in particular have proliferated over the past 18 months. Although it is encouraging to see a greater understanding of the importance of ethics, there is little proof of effective ethical governance practices. Principles are not practice. One must be sceptical of any claims made by organisations unless they, for example, publish the membership and terms of reference of ethics boards along with proof of good ethical practice, since transparency is a fundamental tenet of ethical governance. The five pillars of good ethical governance that

we propose in this paper are intended to close the gap between principles and practice, which is a major theme of this paper.

This is how the paper is organised. In §2, we construct a road map to demonstrate the interconnectedness of ethical governance's constituent elements, such as standards, regulation, responsible innovation, and ethical principles. The roadmap is then discussed in §3, which takes into account public concerns, standards and regulations, safety-critical AI, transparency, and moral machines. In §4, a succinct conclusion is provided, along with a list of five ethical governance recommendations.

Constructing the Roadmap

Our roadmap's central component links ethical research to new regulations and standards. Standards frequently formalise ethical principles into a framework that can be used to assess the degree of compliance or, perhaps more effectively for ethical standards, to give designers instructions on how to lessen the possibility that their product or service will cause ethical harms. Therefore, standards may be explicitly or implicitly based on ethical principles. Examine safety regulations like ISO 13482 [2], where the fundamental ethical tenet is that personal care robots ought to be secure. While many standards do not explicitly state this principle, ISO 13482 does. For example, it could be argued that process standards like the ISO 9000 family of quality management standards embody the idea that everyone benefits from shared best practices. However, regulations that require systems to be certified as compliant with standards or portions of standards are sometimes necessary. The majority of standards are optional. Adopting IEEE 802.11 (WiFi), for example, in a new networked product is not required, but it would obviously be a bad business decision to do otherwise. Furthermore, because a licence to operate a system would not be issued until the system has been demonstrated to comply with the required standards, those standards—often related to safety—are de facto directed. Furthermore, governments can and do influence and direct the adoption of standards—across an entire supply chain—without explicit regulation by making adherence to standards a requirement for awarding procurement contracts. This is another way that soft governance contributes to the adoption of standards. We contend that ethics (or ethical principles) lead to standards, which in turn lead to regulation, as illustrated in figure 1, and that this characterisation has value in comprehending the terrain of ethical governance, acknowledging that this is a simplification of a process with numerous intervening factors.

A few fundamental ethical frameworks are cited in Figure 1, such as the EPSRC Principles of Robotics [4] and the 2006 EURON Roboethics Roadmap [3]. Ten distinct sets of ethical principles, including Asimov's Laws of Robotics1, had been proposed by December 2017; seven of these appeared in 2017, according to an informal survey conducted at the end of 2017 [5]. Table 1 contains a list of these.

These principles share many similarities, most notably that IAS should (i) do no harm, including being devoid of prejudice and deceit; (ii) respect human rights and freedoms, including privacy and dignity, while fostering well-being; and (iii) be transparent and trustworthy while maintaining the locus of accountability and responsibility with their

human designers or operators. The most significant finding may not be related to the content of these principles, but rather to the fact that they are being published more frequently, which is unmistakable proof of the growing recognition of the critical need for ethical principles for IAS. However, principles are not the same as practice. Although they are only the first step, they are a crucial and essential foundation for ethical governance.

Recent standards like ISO 13482 [1] (Safety requirements for personal care robots) and BS 8611:2016 [18], which may be the first ethical robotics standard in history, are cited in figure 1. While BS 8611 covers all classes and domains of robots and robotic systems, ISO 13482 focusses on personal care robots.

Both standards and ethics are part of a larger, more comprehensive framework for responsible research and innovation (RI). Over a ten-year period, RI initiatives in academia, policy, and legislation emerged.

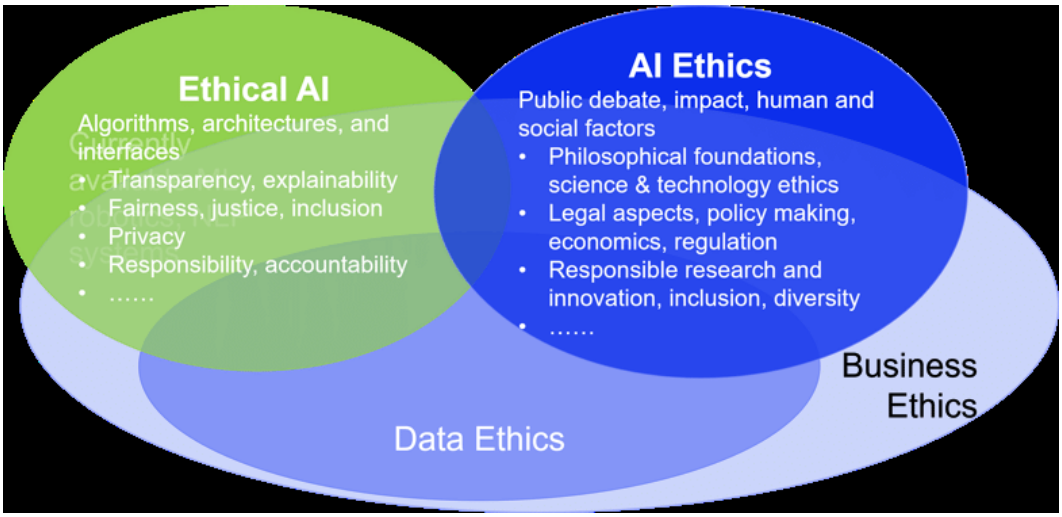

Figure 1. Connecting regulations, standards, and ethics.

The intention of identifying and addressing the risks and uncertainties related to new scientific fields. In recent years, this has broadened to encompass informatics, computer science, robotics, and information and communications technology (ICT) in general. A novel approach to research and innovation governance is put forth by RI [9]. By including strategies for promoting more democratic decision-making through increased inclusion of broader stakeholder communities that might be directly impacted by the introduction of novel technologies, the goal is to guarantee that science and innovation are carried out in the public interest.

Figure 2 illustrates how ethics and standards are both informed by and supported by responsible innovation. Crucially, one of the main tenets of RI is ethical governance. Public engagement, open science, and inclusivity are just a few examples of how RI directly relates to ethics; open science has been referred to as a "trust technology" [2]. The capacity to transparently and methodically measure and compare system capabilities, usually using benchmarks or standardised tests, is another essential element of RI [8].

Another crucial component of RI is the requirement for verification and validation to guarantee both safety and suitability for use, particularly when systems enter real-world applications. For safety-critical systems, compliance with published standards may be a legal requirement that must be met in order for the system to be certified. Verification and validation may be conducted against these standards. Therefore, standards and regulations are linked to validation and verification. The 2014 Rome Declaration on Responsible Research and Innovation, the EPSRC Anticipate, Reflect, Engage and Act (AREA) framework, and the newly founded Foundation for Responsible Robotics [4] are some of the foundational frameworks for responsible innovation that are cited in Figure 2. Additionally, the AREA framework has been customised for ICT in particular [6]. Generally speaking, people trust technology if it is beneficial, safe, well-regulated, and, in the event of an accident, thoroughly investigated.

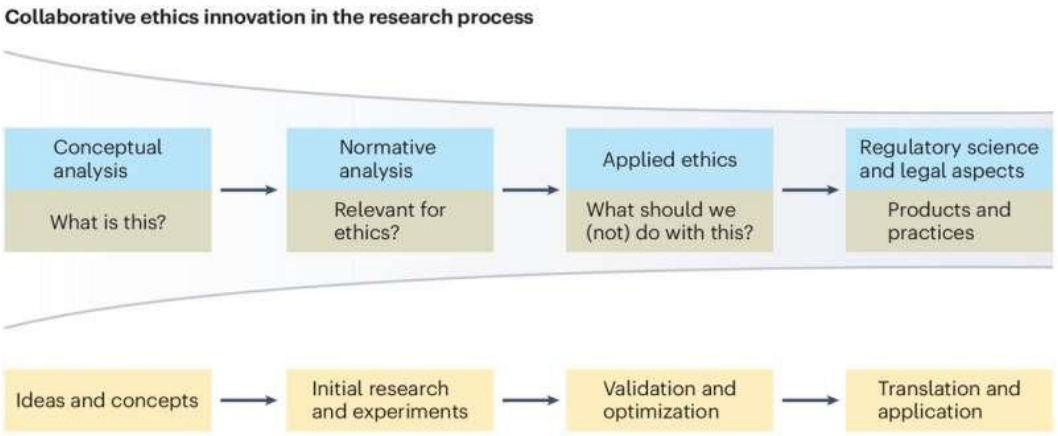

Figure 2. Supported by ethical innovation and research

For instance, we are aware that they operate in a highly regulated sector with a stellar safety record. Good design is not the only factor contributing to commercial aircraft's high level of safety; rigorous safety certification procedures and strong, publicly accessible air accident investigation procedures are also important. It makes sense to propose that some robot types, like driverless cars, should be governed by an organisation akin to the Civil Aviation Authority (CAA), with the Air Accident Investigation Branch serving as the analogue for driverless cars. It is crucial to remember that air accident investigations are social reconstruction processes that must be seen as fair and strong. They also act as a kind of closure to prevent aviation from becoming permanently stigmatised in the eyes of the public. We expect robot accident investigations to play very similar roles.

In order to provide transparency and confidence in the strength of regulatory processes, regulation necessitates regulatory bodies that are connected to public engagement. which, as figure 3 illustrates, all aid in the process of fostering public trust.

Analysis of the IAS Ethics Roadmap

We offer a more thorough analysis of several aspects of the roadmap discussed above, such as public concerns, standards, and regulations, in the sections that follow. With an introduction to safety-critical AI, the necessity of transparency, and—looking ahead—the governance concerns of moral machines (systems that explicitly reason about ethics), we further expand and deepen the roadmap's present and future context.

(a) Public fears

It is well understood that there are public fears around robotics and artificial intelligence. While many of these concerns are clearly unfounded and may have been stoked by media and press hype, some are based on real concerns about the potential effects of technology on things like jobs and privacy. The most recent Eurobarometer survey on autonomous systems showed that the proportion of respondents with an overall positive attitude has declined from 70% in the 2012 survey [1] to 64% in 2014 [2]. Notably, the 2014 survey showed that the more personal experience people have with robots, the more favourably they tend to think of them; 82% of respondents have a positive view of robots if they have experience with them, whereas only 60% of respondents have a positive view if they lack robot experience. It's also noteworthy that a sizable majority (89%) think that autonomous systems are a form of technology that requires careful management.

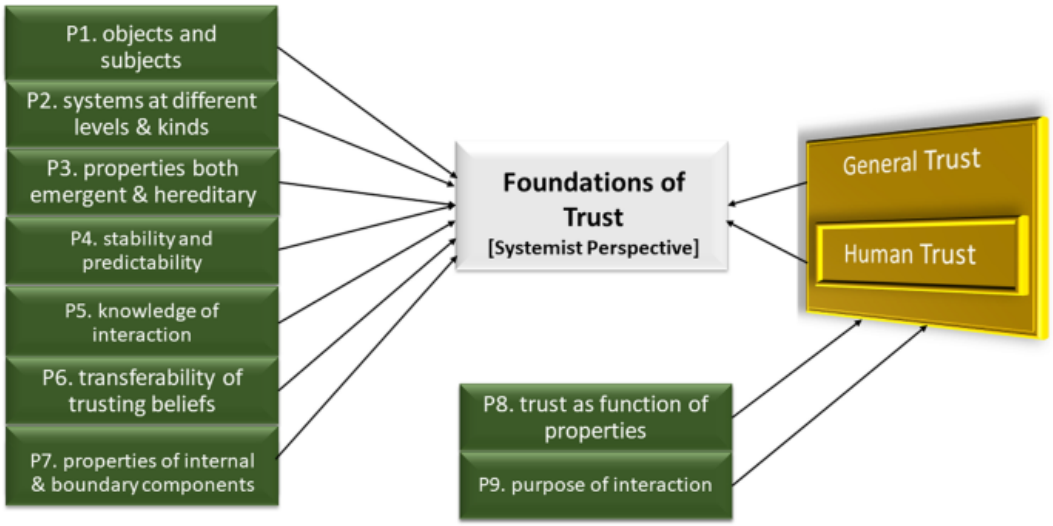

Figure 3. Building public trust.

A recent survey of decision-making in driverless cars reveals distinctly ambivalent attitudes: '... participants approved of utilitarian Autonomous Vehicles (AVs) (that is, AVs that sacrifice their passengers for the greater good) and would like others to buy them, but they would themselves prefer to ride in AVs that protect their passengers at all costs. The study participants disapprove of enforcing utilitarian regulations for AVs and would be less willing to buy such an AV' [30]. It is clear that public trust in IAS cannot simply be assumed [31–33]; to do so could risk the kind of public rejection of a new technology seen

(in Europe) with genetically modified foods in the 1990s [3]. Proactive actions to build public trust are needed, including, for example, the creation of a 'machine intelligence commission' as argued by [6]; such a commission would lead public debates, identify risks and make recommendations to Parliament, for new regulation or regulatory bodies, for instance, and recommend independent mechanisms for responsible disclosure.

Regulations and standards

BS 8611: Guide to the ethical design and application of robots and robotic systems was published in April 2016 as a result of the work of the British Standards Institution Technical Subcommittee on Robots and Robotic Devices [8]. The EPSRC Principles of Robotics are incorporated into BS 8611, which is not a code of practice but rather offers "guidance on the identification of potential ethical harms and provides guidelines on safe design, protective measures and information for the design and application of robots." A wide range of ethical hazards and their mitigation are described in BS 8611, including risks related to society, applications, commerce and finance, and the environment. It also gives designers instructions on how to evaluate and subsequently lower the risks associated with these ethical hazards. Loss of trust, dishonesty, confidentiality and privacy, addiction, and unemployment are some of the societal risks.

Ethically aligned design (EAD), currently in its second iteration, is the main product of the IEEE Standards Association's global ethics initiative. The work of 13 committees within EAD includes: how to embed values; general (ethical) principles into autonomous intelligent systems; techniques to direct ethical design and design; safety and benefits of artificial general intelligence and artificial superintelligence; personal data and individual access control; redefining autonomous weapons systems; humanitarian and economic concerns; law; affective computing; traditional ethics in AI; policy; mixed reality; and well-being. A list of more than 100 ethical concerns and suggestions is presented by EAD. Every committee was asked to suggest problems that ought to be resolved by a new standard. As of this writing, 14 standards working groups are creating candidate standards, or so-called IEEE P7000 "human" standards, to address an ethical issue raised by one or more of the 13 committees listed in EAD. For instance, IEEE P7001 Transparency in Autonomous Systems is developing a set of quantifiable, testable transparency levels for each of the various stakeholder groups, such as accident investigators, certification bodies, and users.

The EU project RoboLaw recently conducted important work on surveying the state of robotics regulation. A thorough report titled Guidelines on regulating robotics is the project's main product. Both ethical and legal aspects are reviewed in that report; the legal analysis includes rights, insurance and liability, privacy, and legal capacity. "The field of robotics is too broad, and the range of legislative domains affected by robotics is too wide, to be able to say that robotics by and large can be accommodated within existing legal frameworks or rather require a lex robotica," the report's conclusion states, focussing on driverless cars, surgical robots, robot prostheses, and care robots. While robotics can probably be well regulated by cleverly adapting existing laws, for some application types

and regulatory domains it may be helpful to think about developing new, fine-grained regulations that are specifically tailored to the robotics at issue.

Artificial Intelligence that Prioritises Safety

Robotics was the focus of early ethics and regulation efforts; AI ethics have only recently come into focus. Compared to distributed or cloud-based AIs, robots are definitely easier to define and regulate because they are physical artefacts. This, along with the already widespread use of AI (in search engines, machine translation systems, or intelligent personal assistants, for instance), strongly indicate that more urgent attention must be paid to the ethical and societal implications of AI, including its governance and regulation.

"An embodied AI" is a fair way to describe a modern robot [38]. Therefore, when thinking about robot safety, we also need to think about the AI that is in charge of the robot. An embedded artificial intelligence (AI) of a suitable level of sophistication will control the three types of robots shown in Figure 1: drones, autonomous vehicles, and assistive robots. However, all of these systems are safety-critical, meaning that their safety is essentially dependent on the AIs that are embedded in them. The decisions that these AIs make have a real impact on human safety or well-being because a failure could result in serious harm or injury. Let's look at two broad concerns with AI: (i) transparency and trust, and (ii) validation and verification. These concerns are particularly relevant to our three exemplar robot categories: drones, autonomous vehicles, and assisted-living robots.

AI Systems Seriously Cast Doubt on Transparency and Trust:

In general, how can the public have faith in the application of AI systems in decision-making, and how can we trust the decisions made by an IAS?

How do we look into the reasoning behind an IAS's decision if it turns out to be catastrophically incorrect, and who is at fault (keeping in mind that the AI cannot be held accountable)?

Safety-critical systems that are currently in place are neither AI systems nor do they integrate AI systems. The reason is that many people believe that it is impossible to verify AI systems—and especially machine learning systems—for applications that are crucial to safety. It is necessary to comprehend the causes of this.

The first issue is the verification of learning systems. Since a learning system by definition changes its behaviour, current verification techniques usually assume that the system being verified will never change its behaviour. As a result, any verification is likely to be deemed invalid once the system has learnt.

The black box problem comes in second. Artificial neural networks (ANNs) are the foundation of contemporary AI systems, particularly the ones that are getting the most attention—so-called deep learning systems. When an ANN is trained using datasets, it becomes nearly impossible to analyse its internal structure to determine how and why it makes a given decision. This is one of the ANN's characteristics. An ANN's decision-making process is opaque.

Although it may not be an intractable problem, the verification and validation of learning systems is currently being studied; for instance, work on autonomous system verification and validation. Although ANNs may not be able to solve the black box problem, algorithmic approaches to AI—that is, those that do not employ ANNs—may be able to. Interestingly, a recent report suggested that "core public agencies... Stop using algorithmic systems and "black box" AI.

Openness

"Transparency" is one facet of ethical governance that was covered above. It would be difficult to argue that opaque governance is ethical; transparency is a necessary component of ethical governance. Both process and product transparency should ideally be demonstrated by ethical governance in robotics and AI; the former refers to the transparency of human research and innovation processes, while the latter refers to the transparency of the robot or AI systems thus developed.

Now think about product transparency. This will inevitably mean different things to different stakeholders—it is obvious that the types and degrees of transparency needed by an accident investigator or safety certification body will differ from those needed by the user or operator of the system. Systems should ideally be transparent to experts and explainable, or even able to explain their own actions to non-experts. The body of research on transparency is expanding; examples include studies on explainability and transparency in robot systems, transparency and the EU General Data Protection Regulation (GDPR), and transparency's limitations.

Finding out why an autonomous system made a certain decision should always be possible (especially if that decision has caused or might cause harm). This is a crucial underlying principle. Since several fatal accidents have already been caused by real-world trials of driverless car autopilots, transparency is obviously urgently needed to determine how and why those accidents happened, address any operational or technical issues, and establish accountability. In order to "provide a guide for self-assessing transparency during development and suggest mechanisms for improving transparency," a new IEEE standard, P7001 Transparency in Autonomous Systems, is presently being developed.

The equivalent of an aircraft flight data recorder (FDR) would be a technological advancement that would offer such transparency, particularly to accident investigators. Because aircraft FDRs are frequently referred to as "black boxes"5 and because such a device would be a crucial physical component supporting the ethical governance of IAS, we call this an ethical black box (EBB). The EBB would continuously record sensor data, just like its aviation counterpart pertinent internal status data in order to significantly aid (though not ensure) the process of determining the reason behind a specific decision or sequence of decisions made by a robot or artificial intelligence, particularly those preceding an accident. Although it is likely that each application domain would have a different standard—one for driverless cars, another for drones, and so forth—EBBs would still need to be designed and certified in accordance with industry-wide standards.

Moving Towards Ethical Machines

Rather than ethical robots, the main focus of this paper is robot and AI ethics. However, autonomous systems in the near future—most notably, self-driving cars—are inherently moral agents. Even if assistive (i.e., care) robots are not specifically built to incorporate ethical values and moderate their decisions in accordance with those values, it is evident that both autonomous vehicles and these robots make morally significant decisions. The values of their creators or, more concerningly, training datasets are arguably implicitly reflected in all autonomous systems (as is evident in AI systems that exhibit human biases. A helpful distinction between explicit ethical agents, which are machines that either directly encode or learn ethics and make decisions based on those ethics, and implicit ethical agents, which are machines built to avoid unethical outcomes. A major outcome of ethical governance is that all robots and AIs should be built as implicit ethical agents. There is growing agreement that near-future robots will, at the very least, need to be designed to reflect the ethical and cultural norms of their users and societies.

The next logical (though technically extremely difficult) step is to give intelligent systems an ethical governor in addition to incorporating values into their design. That is, a procedure that enables a robot or artificial intelligence (AI) to assess the effects of its (or others') actions and adjust its own behaviour in accordance with a set of ethical guidelines.6 The development of workable ethical governors is still the focus of fundamental research and poses two major challenges: (i) the philosophical problem of formalising ethics in a way that makes it easy for machines to implement, and (ii) the engineering problem of implementing moral reasoning in autonomous systems.

Conclusion

Although there is no lack of good ethical principles in robotics and artificial intelligence, we have argued in this paper that there is little proof that these principles have yet to be implemented in the form of efficient and open ethical governance. Naturally, ethical behaviour begins with the individual. Other ethical systems than consequentialism should serve as the foundation for ethical agents; however, computationally modelling such systems is still a challenging research issue and new codes of ethics for professionals, like the recently released ACM code, are highly encouraging. However, strong institutional frameworks and moral leadership are necessary to empower and support individuals. What can we anticipate from AI and robotics firms or associations that profess to follow ethical governance? The ethical governance of AI and robotics has been examined in this paper. Although there isn't a single answer, the paper makes the case that ethical governance will be essential to fostering public confidence in robotics and AI. Without transparent, inclusive, and flexible ethical governance by the companies that create and run them, it is difficult to imagine how disruptive new IAS technologies—like driverless cars, assistive robots, or medical diagnosis AIs—will be widely accepted and trusted.

References

[1] Nersu, S., S. Kathram, and N. Mandaloju. (2020) Cybersecurity Challenges in Data Integration: A Case Study of ETL Pipelines. Revista de Inteligencia Artificial en Medicina. 11(1): 422-439.

[2] Mandaloju, N. kumar Karne, V., Srinivas, N., & Nadimpalli, SV (2021). Overcoming Challenges in Salesforce Lightning Testing with AI Solutions. ESP Journal of Engineering & Technology Advancements (ESP-JETA). 1(1): 228-238.

[3] Gudepu, B.K. and O. Gellago. (2018) Data Profiling, The First Step Toward Achieving High Data Quality. International Journal of Modern Computing. 1(1): 38-50.

[4] Jaladi, D.S. and S. Vutla. (2017) Harnessing the Potential of Artificial Intelligence and Big Data in Healthcare. The Computertech. 31-39.

[5] Pasham, S.D. (2019) Energy-Efficient Task Scheduling in Distributed Edge Networks Using Reinforcement Learning. The Computertech. 1-23.

[6] Jaladi, D.S. and S. Vutla. (2018) The Use of AI and Big Data in Health Care. The Computertech. 45-53.

[7] Gudepu, B.K. (2016) The Foundation of Data-Driven Decisions: Why Data Quality Matters. The Computertech. 1-5.

[8] Pasham, S.D. (2017) AI-Driven Cloud Cost Optimization for Small and Medium Enterprises (SMEs). The Computertech. 1-24.

[9] Pasham, S.D. (2018) Dynamic Resource Provisioning in Cloud Environments Using Predictive Analytics. The Computertech. 1-28.