Ethics in AI and Autonomous System Applications Design

THERE seems to be no end to the development of principles and guidelines in the artificial intelligence/autonomous systems (AI/AS) discipline. Institutes, corporations, governments, each with their imperatives, are rushing to make claims about their practices. In fact, what we have today is a plethora of idealistic design principles that draw from the domain of ethics, that at times the very organizations and agencies that created them struggle to put into practice. This editorial provides examples of AI/AS applications and service offerings, examining the potential for embedding ethics in the design process to minimize end-user vulnerability.

I. Machine Ethics

We are bombarded with media hype about hopes for, and harms of, intelligent and complex systems, big data analytics [item 1) in the Appendix] and machine learning, robotics, and artificial intelligence [item 2) in the Appendix], hyperautomation, and the human versus machine debate [item 3) in the Appendix]. Yet serious research especially from those with engineering and information and communications technology (ICT) backgrounds, alongside ethicists and end-users has been severely lacking. The hype may be predicting the end of the world as we know it [item 4) in the Appendix], "as autonomous systems make decisions to send in the drones indiscriminately," while others herald a vision of an augmented human existence where sustainability is present in all facets of life, the removal of "heavy lifting" for all individuals, and peace on earth to focus on all the right things through collective awareness [item 5) in the Appendix]. The centrist view admits to a middle way that is neither utopic or dystopic, where all things are possible but not necessarily probable, and where humans might well get it right some of the time, but not all the time [item 6) in the Appendix]. The cautious optimist in AI/AS will be confident regarding the future of machines (hard or soft), but will be ready, if not expectant, that there will be difficulties, failures, and even human rights abuses along the way [item 7) in the Appendix].

Despite the discussion and speculation surrounding the topic of machine ethics, ranging from "how can machines have ethics when they have no cognition?" right through to "what would it mean for AI to have a soul," [item 8) in the Appendix] foremost in our minds should be the word "artificial" that precedes "intelligence." We are not looking deep into the machine with anthropomorphic hopes as if it has somehow acquired the "breath of life," but rather understanding it as an entity deliberately designed and implemented by humans that runs using a set of instructions. The instructions can be supervised or unsupervised, they can be case-based or case-free, a machine may be steered or learn by its own trial and error, but independent of the approach, a human has intervened somewhere to set the machine on its trajectory. While embedded software embodied in machines is capable of independently influencing a swarm of microscopic drones, for instance, each microscopic drone possessing Octopus-like capabilities (eight arms independently working through eight axial nerve cords connected to its brain), they are still programmed to perform a given set of possible actions. A machine programmed to sort objects, will not simply break an object on its own volition, although it might learn a best sequence for sorting that has not been preprogrammed.

As stakeholders have come together to work on advancing robotics through a multiplicity of integrated tools, sensors, and software, we have moved beyond the consolidation of the information age to a declaration that we are at the cusp of a Fifth Industrial Revolution. We have witnessed some significant historical AI moments where expert humans have been beaten by machines in Chess (IBM's Deep Blue), Jeopardy (IBM's Watson), and Go (Google's DeepMind) among others [item 9) in the Appendix], where machines seemingly have played or argued more comprehensively and convincingly than humans [item 10) in the Appendix]. Yet a near exhaustive lexicon of definitions, examples, socio-technical relationships, cultural digests, languages, geographic boundaries, jurisdiction-based case law, and even the stealth of a physical robot that is faster, larger, and can carry more, still cannot reckon with human metaphysics [item 11) in the Appendix]. IBM's DeepBlue and Google's DeepMind might well be great at strategizing to beat a human in a board game, and even IBM's Project Debater in its early phases of development might well outstrip a human debater, but put simply, a computer cannot love [item 12) in the Appendix], and they do not pray [item 13) in the Appendix]. We must be cautious not to rush to delegate complex tasks to "intelligent" machines over the human faculties of reason, memory, perception, will, intuition, and imagination that are far superior to things that humans themselves have built. Senses cannot be replicated by sensors, no matter how smart they are. Biomimetic machines, for example, will always be "imitations" of models, systems, and elements of nature for the purpose of solving complex human problems, but they will never be the real thing. This does not mean, we cannot collaborate with the artificial and imitations, to augment our intelligence, but we should never think we can replace it altogether, like a spare part "jacked in" [item 14) in the Appendix]. The sci-fi nightmare of "autonomous

Digital Object Identifier 10.1109/TTS.2020.3019595

anthropomorphic life-sized robots with AI on-board" akin to the stealth of dinosaurs that once roamed the earth, should not be on our technological roadmaps for the 22nd century. And life could only get to *War Games* [item 15) in the Appendix] proportions if we based all our fail-safe systems on machine dependencies without a "human in the loop" and called on technocratic decision-making to be the underlying basis for what we call "government" [item 16) in the Appendix].

II. Current AI/AS Product/Process Offerings

Some treat AI/AS as a futuristic technology but there are many different types of AI/AS capable products that are sold today in major retail outlets for consumer, personal, and household use, categorized as social robotics. On the organizational/industrial side, AI/AS has long been experimented with to reduce operational costs and increase performance at order fulfilment centres (e.g., picking order process) [item 17) in the Appendix]. Dynamic mobile fulfilment robots now vastly outnumber human operators in these packaging centres. Defence-related uses of AI/AS have been available in different military applications, such as in unmanned aerial vehicles (UAVs) or remotely piloted aerial systems (RPASs). The first targeted killing using an unnamed Predator drone was purported to be in Afghanistan in 2002 actioned by the U.S. Central Intelligence Agency (CIA). Table I provides a representative list of AI/AS product offerings. Some of the issues around emerging AI/AS technologies relate to how they will be used in existing contexts and their commensurate social implications.

In the consumer space, we can point to Mattel's Wi-Fi-enabled "Hello Barbie" and Genesis Toys' "My Friend Cayla" doll as AI-based toys for children. In the consumer household market segment, by far, the most successful AI/AS has been iRobot's Roomba vacuum cleaner. By 2014, iRobot had shipped more than ten million Roomba robotic vacuum cleaners, by 2018 they had sold more than 20 million worldwide and by 2020, 30 million robots. The home seems to be a ripe environment for robotics, as a means to penetrate household human activity, monitoring profiles with advanced concept designs in indoor mapping and navigation. The introduction of Google's NEST device at the heart of home safety systems, alongside Google's Dropcam are two more AI-based technologies. Additionally, devices, such as Amazon's Echo, Amazon's Ring, and Google's Home, integrated with Philips Hue Lighting capabilities, are really leading in the home automation and convenience spaces.

Chatbots, such as LivePerson, LiveChat, Amazon Lex, Dialogflow, IBM Watson, and bold360, are other examples of disembodied AI, as well as personified Personal Assistants (also known as hydras) like Apple Siri, Microsoft Cortana, Amazon Alexa, and Google Now. Other bots include Microsoft's ill-fated Tay, Talking Angela, Cyberlover, and Microsoft's Chinese Xiaoice. A controversial space now covered by some smaller commercial offerings includes sexbots, life-sized anthropomorphic objects attempting to provide companionship, and/or sexual stimulation. Redirecting attention to such inventions detracts from the needs of those living with disabilities that could be solved through AI/AS innovation.

In the organizational sector, we have witnessed aged care facilities adopt carebots for their clients, inclusive PARO the robotic harp seal and other robots for the movement of linen, and other repetitive tasks [item 18), 19) in the Appendix]. A variety of anthropomorphic robots have graced museums, such as SoftBank Robotics' Pepper, Honda's ASIMO, and Sony's AIBO and QRIO to name a few. All of these help us to see the evolution toward social robotics, socially intelligent machines that may seamlessly become integrated into our everyday life [item 20) in the Appendix]. KnightScope's K5 made headlines in 2014 by unveiling a ground vehicle drone for security purposes, trialled at a variety of high-profile HQs, including that of Microsoft [item 21) in the Appendix]. Boston Dynamics has repeatedly stunned the world with their LS3, Cheetah, ATLAS, and PetMan inventions which seemingly have military application [item 22) in the Appendix]. There are additional emergency services-based AI, especially in locational analytics, and in the defence sector with respect to drones.

Fig. 1. Illustrations of current AI/AS/robotics offerings.

III. Ethical Autonomous Systems

Perhaps this nascent stage of the Fifth Industrial Revolution is about varied stakeholders in society (industry, government, academia, and citizenry) gaining a grasp of what machine ethics actually is. Using the terminology of James Moor [item 23) in the Appendix], it is important to refer to both implicit ethical agents, that is, machines designed to avoid unethical outcomes, and explicit ethical agents, that is, machines which either explicitly encode or learn ethics and determine actions based on those ethics. Of course, ethical machines are socio-technical cyber-physical systems [item 24) in the Appendix], so that exploring the educational, societal, and regulatory implications of machine ethics, including the issue of ethical governance, is also needed. Ethical governance is needed to develop standards and processes that allow us to transparently and robustly assure the safety of ethical autonomous systems and hence build public trust and confidence. Moor identifies [item 24) in the Appendix] four different categories of ethical agency: 1) ethical impact agents (a system that can be evaluated); 2) implicit ethical agents (constrained to avoid unethical outcomes); 3) explicit ethical agents (able to reason about ethics); and 4) full ethical agents (able to justify judgements). This suggests a "bright line" between the last two and, though an AI might never cross that line, Winfield argues his Asimovian example qualifies as an explicit ethical agent [item 25) in the Appendix]. Table II presents an AI/AS typology matrix.

Machines can come in three forms: 1) hardware-based; 2) software-based; or 3) hybrids. Hardware-based machines increasingly have software components driving decisions. Software-based machines may be disembodied by any physical form, save for a series of instructions. These have been labeled bots. Bots unlike robots are not encapsulated in anthropomorphic designs, but utilize advanced analytical techniques

TABLE I
CURRENT AI/AS PRODUCT/PROCESS OFFERINGS

AI/AS Context	Example	Description
Consumer	Personal Use	AI email clients for categorisation
	Sex robots	Robots predominantly for male use
	Home Use	Vacuum cleaners (iRobot's Roomba)
	Children's Toys	Interactive toys (Hello Barbie), learning technologies
	Voice Activated Clients	Search engines, keyless entry, preferences
	Home automation Clients	Mood, color, music, convenience (IOT-based)
	Child care	Especially children with autism
	Companionship	Robots for aged care clients to keep interaction
	Health and Fitness	Robots that can guide instruction
	End of Life Robots	Robots for care
Organisational/ Industrial	Industrial Use	Robots for sorting (Kiva Systems)
	Aged Care Facilities	Robots that help with human duties
	Law bots	Case law research
	Online chat bots	To assist in getting information on products or services to human users who need it and don't wish to interact with a human being
	Medical bots	Medical clinical trial searches for optimal solutions
	Social media monitoring	Positive messaging, brand raising, sentiment analysis
	Political bots	Bots for the promotion of political orientation
Emergency Service Organisations	Social media bots (police)	Bots for social media monitoring, prevention of suicide and raising mental health alerts, cyberbullying and more
	Emergency Response Use (emergency)	AI shortest path routing among other locative analytics
Defense	Killer drones	Predator Drones (new microscopic)
	Carrying payload drones	Land-based autonomous systems (LS3 Boston Dynamics drone)
	Surveillance and reconnaissance drones	Watching in the skies 24x7 for enemy attack

like conversational analysis, without even the appearance of an online avatar. Some of these software-based bots have been said to pass the Turing test [item 26) in the Appendix], unbeknownst to human users. There are hybrid machines that take physical form and are ingested with artificial intelligence capabilities that can seemingly respond in intelligent ways. Anthropomorphized social robots that are built out of materials, such as "frubber" (short for flesh rubber), plastic, electronic eyes, and human hair, may be tremendously realistic for theme-park animatronics or science-fiction movie actroids but in reality, have no place in useful robotics. Among these, we could identify Hanson Robotics' Sophia who gained alleged citizenship in Saudi Arabia (even before women had been given the right to drive in the country), or even Boston Dynamics' Atlas or PetMan, or Martine Rothblatt's Bina48.

Fig. 1. Top left: Paro therapeutic robot seal (Aaron Biggs, 2005), Bottom Left: Le robot NAO de l'entreprise Aldebaran Robotics au salon Innorobo à Lyon en 2015 (Xavier Caré, 2015). Top Middle Left: iRobot Create with mounted camera and minicomputer (Jeremiah, 2007). Top Middle Right: Actroid (Jennifer, 2006). Bottom Middle Left: Koromoroid, Child Android (Franklin Heijnen, 2015). Bottom Middle Right: Charge of the fembots, Robot Restaurant, Shinjuku, Tokyo, Japan (Cory Doctorow, 2014). Right: Robot Atlas (Kansas City News).

TABLE II
AI/AS TYPOLOGY MATRIX

	Hardware	Software	Hybrid
Manual	A system that can be evaluated.		
Some Artificial Intelligence	A system that can be evaluated or constrained to avoid unethical outcomes.		
Semi-Autonomous	A system that can be constrained to avoid unethical outcomes or be able to reason about ethics.		
Fully Autonomous	A system that can justify judgements.		

Machines that are imbued with ethics may receive instructions manually, partially through AI, semi-autonomously or autonomously in nature. They may be hard/soft-ware based or hybrids.

IV. CYBERETHICS: EMBRACING ETHICS IN DESIGN AND DEVELOPMENT AND EMBEDDING ETHICS IN THE MACHINE

Ethics in ICT and engineering can pertain to:

1) whether or not a system is ethical to begin with in a technically deterministic way (e.g., killer drones) and whether or not they uphold universal ethical principles, such as those noted in the Universal Declaration of Human Rights in the social shaping of technology [item 27) in the Appendix];
2) whether a system is designed using methodologies and development cycles that incorporate ethics (also known as value-driven design and alignment) [item 28) in the Appendix];
3) whether or not an organization has ethics-based training for its employees and incorporates ethics in its vision and mission statements with an overseeing governing body;
4) whether an individual supporting the development of a system has sufficient ethical training (e.g., are they abiding by codes of conduct and codes of ethics as practitioners, say for instance, as a member of the IEEE or ACM) [item 29) in the Appendix];
5) whether or not an end user (mis)applies an existing system in an (un)ethical way.

A. Three Levels of Ethics in ICT and Engineering

Ethics in ICT and engineering can be applied at three levels: macro, meso, and micro [item 30) in the Appendix]. Macro ethics are the international approaches to oversight, the law, and high-level regulations governing the development, production, and distribution of intelligent machines. Macro ethics also involves nongovernment organizations (NGOs) that observe and report on current happenings. Some examples include the Rome Declaration on Responsible Research and Innovation and EPSRC/AHRC Principles of 2010 but also international standards, such as ISO 13482:2014—Robots and robotic devices—safety requirements for personal care robots [item 31) in the Appendix], and international NGOs like the International Committee for Robot Arms Control (ICRAC).

TABLE III
REPRESENTATIVE MACRO, MESO, AND MICRO ETHICS BODIES, ORGANIZATIONS AND SOCIETIES RELATED TO AI/AS

Ethics Level	Law, Convention, Principles, Framework, Alliance, or NGOs Relevant to AI/AS
Macro	☐ Universal Declaration of Human Rights 1948 ☐ International Human Rights Convention ☐ International Covenant on Civil and Political Rights ☐ Convention on Cybercrime ☐ Rome Declaration on Responsible Research and Innovation ☐ EPSRC/AHRC Principles of 2010 ☐ RoboEthics Roadmap 2006 ☐ ISO 13482:2014 - Robots and robotic devices -- Safety requirements for personal care robots ☐ Open Science Reproducibility ☐ American Civil Liberties Union ☐ Human Rights Watch ☐ Amnesty International ☐ International Committee for Robot Arms Control ☐ Foundation for Responsible Robotics ☐ Campaign Against Sex Robots
Meso	☐ The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems - Ethically Aligned Design Standards P7000 Series ☐ BS 8611:2016 - Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems ☐ Partnership on Artificial Intelligence to Benefit People and Society ☐ SANS Institute (Organisation level policies)
Micro	☐ Association for Computing Machinery (ACM Code of Ethics and Professional Conduct) ☐ Australian Computer Society (ACS Code of Ethics and ACS Code of Professional Conduct) ☐ British Computer Society (BCS Code of Conduct) ☐ Computer Ethics Institute (Ten Commandments of Computer Ethics) ☐ IEEE (IEEE Code of Ethics and IEEE Code of Conduct) ☐ League of Professional System Administrators (The System Administrators' Code of Ethics)

Meso ethics are organizational centric-approaches and at times are led by industry guidelines and professional body standards bringing stakeholders of all types together in the autonomous systems value chains. These are characterized by professional industry standards, such as the IEEE's Ethically Aligned Design statement P7000 series [item 32) in the Appendix], bringing together practitioners and members of professional bodies to designate best practice. Another example of a standard we can point to at the industry level includes BS 8611:2016—Robots and robotic devices—guide to the ethical design and application of robots and robotic systems [item 33) in the Appendix]. In some instances, organizations may also encourage value-driven design by designating personnel to such specific areas as human rights and advanced systems (public) policy. There are also internally driven oversight mechanisms forming in private companies, such as the Partnership on Artificial Intelligence to Benefit People and Society that involves the Big Five, Facebook, Amazon, IBM, Microsoft, and Google in an alliance whose primary mission is to share in best practices on transparency and ethics in the development of AI and the education of the general public [item 34) in the Appendix].

Micro ethics are traditionally concerned with individual membership to various not-for-profit organizations. These may include but are not limited to: the IEEE, the ACM, the ACS/BCS, and other societies that incorporate ethical responsibility as central tenets of their members. There are also established ethics hotlines for employees, but these hotlines are generally not targeted to the domain of ICT and Engineering, but more broadly to associations like Whistleblower Societies or Ethi-call services such as those available at Australia's Ethics Centre. Table III provides further examples.

In all the various types of cyber ethics/traditional ethics that stakeholders abide by, in every case, we can say that consequences in the form of enforcement and commensurate

penalties are lacking. No one wishes to hold a stick over an industry ripe with opportunities to innovate, but it is also true that we are entering a time of potential mystery in the integration and convergence of new technological subsystems. No one is claiming that innovation processes should be stifled in any way by ethical judgements in design, but rather that if ethics and values are embraced and infused in the innovation process while a technology is still in its nascent phases of exploration and in fact embedded "in the machine," that our future may well be very bright if we are able to harness and capitalize on the new potentialities in a way that reduces harm to the citizenry (e.g., social, personal, economic, medical, safety, job losses, etc.). That is, technological progress is not necessarily inhibited by adopting ethical and stakeholder-inclusive approaches to design.

With the novelty of new incremental and radical innovations comes the novelty of new approaches to development and design science. One recommendation is that patent applications are submitted with commensurate ethics-based applications, further substantiating the value that a given innovation may grant society and providing transparency in terms of the anticipated risks and outcomes. This practice has been embraced in most University structures in the developed world as a direct response to research abuses in the 20th century, among the most notorious the experiments of Nazi physicians. So why are corporations exempt from this process—as if scientists, inventors, and designers in private companies can be trusted more than academics in public institutions? Possibly, in the future, what will make for a good patent should not be merely novelty, but a product or process that in fact can solve a problem that exists now or is likely to exist in the future and is anticipated to have limited negative impacts on society.

For many engineers and ICT practitioners, despite their training in the domain of ethics while gaining certified degrees, ethical use of a technological system is a societal matter, something that happens as society shapes technology and of concern to the humanities. Still, it is bewildering when practicing engineers have deliberated in private conversation, that: "ethics is not my problem I am an inventor and I do not determine how innovations are applied," "ethics can be left to ethicists, that's not my job," and "I have no control over how consumers use my product." However, it is strongly acknowledged by most ethicists, that technology is seldom neutral, if ever. And the more complex technologies become, made up of many components and subsystems, the more a machine can be said to possess exposures, perhaps possessing algorithmic bias, and be laden with the subjectivity of its designer(s) [item 35) in the Appendix]. How do we overcome these kinds of challenges? How do we convince practitioners that ethics is everyone's problem? That if harnessed appropriately and embedded in the design process, that ethics can be used to create even more robust technologies?

All change—in the way we build things—is local. As a conscientious group of practitioners within small units within organizations, or larger groups as members of society, we can ensure a positive outcome if we choose to consider the values of those around us but also those values that have been identified by those before us [item 36) in the Appendix]. Leadership plays a significant role in the adoption of international and national standards pertaining to robotics safety, risk-driven robotics design, and the enactment of targeted industry alliances that have care and risk minimization as a focal point of their development environment.

One of the initiatives that has helped bring stakeholders together across all levels of ethics has been *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, executive directed by John C. Havens. The first three general principles of Ethically Aligned Design [item 37) in the Appendix], a foundational 250+ whitepaper written by over 700 global experts are as follows.

1) *Human Rights:* A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
2) *Well-Being:* A/IS creators shall adopt increased human well-being as a primary success criterion for development.
3) *Data Agency:* A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.

EAD (https://ethicsinaction.ieee.org/p7000/) inspired the creation of more than 13 standards working groups, including a standard that had already launched. Table IV below provides a summary description for each IEEE standard. The design standard could well have been called value-driven design.

One criticism that has been made of IEEE standards in the AI/AS space is that publishing the P 7000 series of standards is only a part solution. What must follow the working group agreement and drafting and redrafting toward finalization, is the social pressure to acknowledge that industry *must* conform to these standards, despite, for example, expected resistance by autonomous vehicle manufacturers to follow suit. Once the major players conform to the standards granting regulators and the courts something to work with, to enforce the standards, it then becomes a race to compliance and not just mere principles or guidelines.

V. Are Global Ethics the Answer?

To this end the question becomes, can we speak of a universal global ethic? Are there generic rights and wrongs with respect to machine ethics? If we argue against a global ethic, we may, for instance, make the claim that autonomous systems built in the United States may have American values, and those designed in China may have markedly different values based on a different economic system and cultural identity, as opposed to those in nations like Canada, Australia, and New Zealand with relatively small populations and varying levels of geopolitical stealth. What kinds of mechanisms can be introduced to ensure equality, especially in defense-related technologies that utilize artificial intelligence? One nation state may well abide by saying "no" to killer robots, and another may well unleash them as an offset mechanism, a global geopolitical game-changer. Additionally, in the future,

TABLE IV
IEEE ETHICALLY-ALIGNED DESIGN STANDARDS

ID	Name	Working Group Focus
P7000	Model Process for Addressing Ethical Concerns During Systems Design	Outlines an approach for identifying and analyzing potential ethical issues in a system or software program from the onset of the effort. The values-based system design methods addresses ethical considerations at each stage of development to help avoid negative unintended consequences while increasing innovation.
P7001	Transparency of Autonomous Systems	A Standard for developing autonomous technologies that can assess their own actions and help users understand why a technology makes certain decisions in different situations. The project also offers ways to provide transparency and accountability for a system to help guide and improve it, such as incorporating an event data recorder in a self-driving car or accessing data from a device's sensors.
P7002	Data Privacy Process	Specifies how to manage privacy issues for systems or software that collect personal data. It will do so by defining requirements that cover corporate data collection policies and quality assurance. It also includes a use case and data model for organizations developing applications involving personal information. The standard will help designers by providing ways to identify and measure privacy controls in their systems utilizing privacy impact assessments.
P7003	Algorithmic Bias Considerations	Provides developers of algorithms for autonomous or intelligent systems with protocols to avoid negative bias in their code. Bias could include the use of subjective or incorrect interpretations of data like mistaking correlation with causation. The project offers specific steps to take for eliminating issues of negative bias in the creation of algorithms.
P7004	Standard on Child and Student Data Governance	Provides processes and certifications for transparency and accountability for educational institutions that handle data meant to ensure the safety of students. The standard defines how to access, collect, share, and remove data related to children and students in any educational or institutional setting where their information will be access, stored, or shared.
P7005	Standard on Employer Data Governance	Provides guidelines and certifications on storing, protecting, and using employee data in an ethical and transparent way. The project recommends tools and services that help employees make informed decisions with their personal information. The standard will help provide clarity and recommendations both for how employees can share their information in a safe and trusted environment as well as how employers can align with employees in this process while still utilizing information needed for regular work flows.
P7006	Standard on Personal Data AI Agent Working Group	Addresses concerns raised about machines making decisions without human input. This standard hopes to educate government and industry on why it is best to put mechanisms into place to enable the design of systems that will mitigate the ethical concerns when AI systems can organize and share personal information on their own. Designed as a tool to allow any individual to essentially create their own personal "terms and conditions" for their data, the AI Agent will provide a technological tool for individuals to manage and control their identity in the digital and virtual world.

(To be Continued)

with diminishing costs of powerful technologies unconventional warfare using AI/AS may well be possible by nonstate actors with relatively low financial backing.

Scholars are pointing to the usefulness of soft law, as international law and laws of nation states continue to fall short in keeping pace with developments [item 38) in the Appendix]. If we want to develop technologies in an agile fashion, then we need a more agile way of proactively keeping up with the new technologies. We can no longer be reactive and wait for harm to occur, e.g., the death of a human exposed to risks related

TABLE IV
CONTINUED

P7007	Ontological Standard for Ethically Driven Robotics and Automation Systems	Establishes a set of ontologies with different abstraction levels that contain concepts, definitions and axioms that are necessary to establish ethically driven methodologies for the design of Robots and Automation Systems.
P7008	Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems	Establishes a delineation of typical nudges (currently in use or that could be created) that contains concepts, functions and benefits necessary to establish and ensure ethically driven methodologies for the design of the robotic, intelligent and autonomous systems that incorporate them. "Nudges" as exhibited by robotic, intelligent or autonomous systems are defined as overt or hidden suggestions or manipulations designed to influence the behavior or emotions of a user.
P7009	Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems	Establishes a practical, technical baseline of specific methodologies and tools for the development, implementation, and use of effective fail-safe mechanisms in autonomous and semi-autonomous systems. The standard includes (but is not limited to): clear procedures for measuring, testing, and certifying a system's ability to fail safely on a scale from weak to strong, and instructions for improvement in the case of unsatisfactory performance. The standard serves as the basis for developers, as well as users and regulators, to design fail-safe mechanisms in a robust, transparent, and accountable manner.
P7010	Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being	Establish wellbeing metrics relating to human factors directly affected by intelligent and autonomous systems and establish a baseline for the types of objective and subjective data these systems should analyze and include (in their programming and functioning) to proactively increase human wellbeing.
P7011	Process of Identifying & Rating the Trust-worthiness of News Sources	The purpose of the standard is to address the negative impacts of the unchecked proliferation of fake news by providing an open system of easy-to-understand ratings. In so doing, it shall assist in the restoration of trust in some purveyors, appropriately discredit other purveyors, provide a disincentive for the publication of fake news, and promote a path of improvement for purveyors wishing to do so.
P7012	Machine Readable Personal Privacy Terms	The purpose of the standard is to provide individuals with means to proffer their own terms respecting personal privacy, in ways that can be read, acknowledged, and agreed to by machines operated by others in the networked world. In a more formal sense, the purpose of the standard is to enable individuals to operate as first parties in agreements with others—mostly companies—operating as second parties.
P7014	Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems	This standard defines a model for ethical considerations and practices in the design, creation and use of empathic technology, incorporating systems that have the capacity to identify, quantify, respond to, or simulate affective states, such as emotions and cognitive states. This includes coverage of 'affective computing', 'emotion Artificial Intelligence' and related fields.

to testing of autonomous vehicles, and simply say that this is a form of collateral damage as a new technology undergoes fine-tuning. This is akin to building monolithic structures like large walls, buildings, and bridges in the early stages of the 20th century and killing thousands as a mere side-effect in the name of progress. It is purported that one million Chinese lost

their lives in building the Great Wall of China, 28 men lost their life building the San Francisco–Oakland Bay Bridge and 16 men lost their life from work-related incidents in building the Sydney Harbor Bridge. Are we prepared to suffer a similar fate in the 21st century? Buildings are static structures, how much more so might autonomous systems put lives at risk? And what can we do to minimize this risk? [item 39) in the Appendix] The claims being made about the reduction of road accidents have merely been a thought experiment to date. Five people have already lost their lives in response to "testing" autonomous vehicles (four Tesla drivers, and a homeless pedestrian hit by an Uber). Though these are considered Level 2 fatalities and only "automated driving," the question remains, how might this figure increase with full-fledged "autonomous vehicles" allowed on public hybrid highways. The Uberization process that is currently placing pressure on the U.S. automotive manufacturers, is not being garnered by all players in the supply chain. Transport workers, especially, are rebelling.

VI. Risk, Exposure, and End-User Vulnerability

In the end, as any instance, when decisions are delegated to algorithms, who is to blame for the actions taken by autonomous systems that cause harm? Can we simplify legal liability when an autonomous system is at fault? For example, can we blame a collision on propagation delay if a network is required in response to a decision, or blame passenger fatalities on an anonymous hack given lax security engineering defenses? Much insight can be gained by well-known cases in the automotive industry that have played out over the last decade in particular, since computers entered vehicular design in a significant way. In some of the more high-profile cases that were settled into the billions of dollars, while technical defects were identified as being "the cause" because that is what is required to win litigation, the real underlying issues were likely company incentive structures that encouraged a poor safety culture. For players in the self-driving car rivalry, such as Uber ATG, taking risks to win the race to autonomy and gain first mover advantage was calculated but backfired. Other automotive companies during the 2000s, rushed to settle lawsuits instead of address computer system design defects, allowing somewhat "known" faults to go unchecked, armed with the knowledge that vehicle recalls would act to cause significant reputational damage, financial losses, and nightmarish repair scenarios.

What is pertinent from automotive cases to date, and will continue as a trend in autonomous vehicles? Is that the entire judicial system is predicated on finding "someone" or "something" to blame for bad outcomes. Yet, if there is no "blame" assignable (e.g., immunity or "no-fault" settlements), we lose the primary mechanism by which we can apply pressure toward improvement. And what we have witnessed over the last twenty years especially, is that it takes a lot to change this lucrative business and there is a significant power imbalance between consumers, industry players, the judicial system, and regulators. As in most complex socio-technical systems, the companies offering the product or service are able to guard their power advantage, over the regulators, and especially over customers. But it is yet unclear how this might turn out for self-driving cars, in context.

TABLE V
Levels of End-User Vulnerability Dependent on AI/AS

		Types of AI/Autonomous Systems		
		Control	Care	Convenience
Levels of End-User Vulnerability	High	X	X	
	Medium	X	X	
	Low		X	X

Machine ethics can take on various levels of invasiveness and risk. These may be categorized as machines that have low levels of risk, intermediate-risk, or high-end risk. While this is a rudimentary scale of understanding autonomous systems, one way to consider them is in relation to the magnitude of human user vulnerability. Machines that malfunction and cause immediate catastrophic consequences to end users are those that are high end in risk. For example, we can consider biomedical innovations like smart pills that administer drug delivery based on sensors in the body [item 40) in the Appendix]. Giving an individual more or less insulin could be fatal. At the same time, machines making decisions in supervisory control and data acquisition (SCADA)-related control network architectures, have the ability to malfunction if semiautonomous or autonomous, causing mass-scale problems with end-users. On the flipside, low levels of risk related to autonomous systems, may be those that have little or no effect on a human life. We can then consider the following matrix in Table V in relation to granting more attention to higher levels of risk, driven by particular types of services that are intelligent and/or autonomous systems.

VII. Control, Care, and Convenience AI-Based Applications and Levels of Invasiveness: In-body, On-body, or External

One of the major areas in which autonomous systems are said to make a great contribution is in those related to medical AI for human well-being. If we place such responsibility upon technologies to sustain life or at least enhance it, we must be ready to introduce commensurate safeguards during the design process of such systems. Advanced autonomous systems must have global privacy and security by design principles embedded [item 41) in the Appendix]. Hacks as we know them today will morph as our innovations become more complex and we witness an explosion in the number of IP-based devices that are wholly unsupervised by external systems. We may well see hackers become predatory by default in their intent, when a security breach—accidental or deliberate—could cause an instant fatality or series of fatalities. In emerging biomedical innovations in the brain that are there to help sustain life, such as brain implantable technologies for Parkinson's disease sufferers or those suffering from major depressive disorder (MDD), a hack to a sensitive component might cause human distress or worse [item 42) in the Appendix]. In effect, this is an example of predatory hacking that has the capacity

TABLE VI
OPERATIONAL SCENARIOS OF AI/AS ON HUMANS AND MACHINES

AI/AS Flow of Information	*In body*	*On-body*	*Out-of-body (external)*
In body	Within Human	Humancentric	Human to machine
On body	Humancentric	Humancentric	Human to machine
Out-of-body (external)	Machine to human	Machine to human	Machine to machine

to affect an individual sufferer with a near fatal blow depending on the circumstances. This has been cleverly named "death by Internet" by researcher Joseph Carvalko [item 43) in the Appendix].

Control applications dispersed among emergency services organizations can also have differing effects and outcomes. For example, we could ponder that autonomous systems in a fire emergency department, might well aid those calling for help, protect firefighters from imminent threat dangers, and minimize potential damage, and financial losses to a business. Semi-autonomous systems, may also help keep the injured alive as paramedics first arrive on the scene after accidents. Smart systems that help assess an individual, at the scene of a crime or on route to a hospital, might save lives, especially, when access to medical histories might be available, thereby understanding preconditions and not wasting valuable time in response or getting in touch with next of kin. But for now at least, we should not overstate the current capabilities. Companies like IBM at the cutting edge of AI systems, have tried to temper our expectations time and time again, as they openly admit to finding it difficult to turn Watson-intelligence for winning Jeopardy into a useable medical AI product to be implemented during patient consultation or for disease discovery [item 44) in the Appendix].

There are also shortcomings of autonomous systems and those artificial intelligence-based algorithms (e.g., Clearview AI) that might well be used to proactively profile individuals against crime-based hotspots, or even identify or ascertain someone's propensity to commit a crime based on socio-economic demographic data, DNA make up, ancestry, social media posts using sentiment analysis and behavioral analytics, and even credit reporting history. Law enforcement software rolled out in order to maintain police records and dossiers on reoffenders might well begin a pattern of typecasting minority groups, and socially sorting individuals based on anything but proven actions, i.e., crimes [item 45) in the Appendix]. Bringing disparate database systems together to predict criminality is the kind of situational awareness spoken of in the context of uberveillance [item 46) in the Appendix]. Such shortcomings of uberveillance may additionally be prevalent in AI/AS: information manipulation, misrepresentation, and misinterpretation [item 47) in the Appendix]. This is especially true when uberveillance dictates surveillance mechanisms that are in-body, such an embedded black-box monitor and tracking recorder, akin to that in an airplane or vehicle [item 48) in the Appendix]. The uses of these uberveillant technologies [item 49) in the Appendix] might well be for the monitoring of vital signs and physiological characteristics, location movement and condition information, or prescribed medicine-taking behavior (e.g., Abilify Pill) [item 50) in the Appendix]. AI/AS systems can also be categorized as being embedded, adorned on the body (i.e., wearable) and communicating with various flows of abstractions as can be seen in Table VI.

When a decision is delegated to autonomous systems (i.e., machine-to-machine) without a human in the loop and without ethical alignment, the decision and flow on outcomes may be fraught with danger if not tested and verified extensively. Additionally, when human-centric machines interact with external machines, or receive inputs from machines, interventions to these transactional flows of information can be compromised. One way of protecting the security of a system is by continually investing in proprietary solutions that are closed and not known to the open market, despite that patents are supposed to reveal the inner workings of the invention. Traditionally, these kinds of developments in products have also been known as "black boxes" in a different sense, that is, their inner workings are to a great degree unknown [item 51) in the Appendix]. The algorithm (i.e., software) in the hardware is protected by a patent, and the nuts and bolts are secret to a given organization who conceived of the unique approach, in order to protect its intellectual property and market advantage. How will this play out in future court cases? Again, we have seen a glimpse of this in the automotive industry, which has required expert witnesses to gain access to raw code, only through standalone computers in a closed setting, and merely with paper and pen in hand, completely inadequate tools to scrutinize and analyse hundreds of thousands of lines of code with a security guard watching over the shoulder. One way to ascertain what is going on in the black box is to publish standards openly and transparently, allowing the industry at large to monitor developments, including NGOs that do not have the resources to gain access to more (Table VII).

But will this solve the problem of the black box? What happens when ethical agents are called upon to simulate decisions based on real-time or near real-time events? Are these openly reproduceable? Are they interpretable and arriving at decisions through processes that humans can understand and trust? And what happens when a system is smart enough to learn in an unsupervised way? Can such results be verified using formal methods? What happens when several black boxes of different types come head to head in a proximate location? item 53) in the Appendix electromagnetic interference (EMI) as an area of study in AI/AS-based products will rise in

TABLE VII
TWO TYPES OF BLACK BOXES

Types of Black Boxes in AI/AS	Implications
Type I: Algorithmic Mystery: A product or process as a black box	A product or process that is patented to protect a company's intellectual property and market advantage. It follows, that the company's code for that given product or process is "dark" and not transparent to competitors or the general public. It is akin to the KFC's secret sauce, and the ingredients in Coca Cola which cannot be replicated exactly to maintain competitive advantage.
Type II: Uberveillance[52]: Embedded surveillance technologies as a black box	This type of "black box" is akin to an airline black box that comes in two parts. For now, it is technology like a flight data recorder (FDR), that comes in the form of an embedded microchip implant tethered to a smart phone. It is possible that a voice and visual data recorder might be worn or lugged in the future, or interface with lampposts, akin to an airline's cockpit voice recorder (CVR) unit.

demand [item 54) in the Appendix]. For now, communications authorities are saying little publicly, but set on an Internet of Things (IoT) trajectory, we may see situations in which one AI/AS-based biomedical device affects another (e.g., the possibility of one brand of insulin pump to influence the reading levels of another brand of insulin pump in two human recipients).

Future innovations like brain implants in the military lead to questions of whether or not AS encroach on human autonomy [item 55) in the Appendix]. Is it possible in the future that AS systems implanted in the brain will override natural cognition? [item 56) in the Appendix] This was once the stuff that dreams were made of, speculated by Jaques Ellul [item 57) in the Appendix] and others before him. Is it possible to have a truly congruent human and machine at the helm as Norbert Wiener questioned in the notion of "cybernetics" taken from the Greek word meaning "steersman?" [item 58) in the Appendix] And is this possibly something other than human or simply a hybrid system that shares equal rights toward decision-making? Transhumanists, in the citizen science space, anticipate a future where they are plugged in [item 59) in the Appendix], and those like Regina Duggan have even pondered on a direct neural brain-to-computer interface to social media platforms like Facebook [item 60) in the Appendix]. We already have significant investment in brainwave technologies, such as the Emotiv skull cap and the Interaxon MUSE band. This is a significant departure from the potential for AI/AS used in exoskeletons to help those with paraplegia walk or for robodogs as assistants for those living with autism or dementia. The cyberinsurance industry is set to grow based on our own uses of AI/AS systems.

VIII. Social Implications of Personal, Industrial, and Defence AI/AS Products

One of the biggest issues in this space is presently how automation in the field will affect human beings [item 61) in the Appendix]. Beyond the matter pertinent to employment, what is this doing to our heads? In the first instance, when we take the natural world around us, or artifacts in the natural world and we attempt to digitize them, we are altering the very state of that which we are studying as it has taken on a different form. Even deeper, automation places incredible pressures on individuals to be digital-like and to delegate decisions to algorithms without fully understanding the consequences of their actions. All manual processes that are repetitive are being automated and so less and less interaction between humans is occurring. This has implications for how people feel about their workplaces—whether or not they feel their skillsets are indeed being made redundant, and whether or not they can keep up with the changes occurring in their workplace. When we do not utilize skills that we have imbued as part of our human function, we feel a type of atrophy occurring both in our abilities and our minds. How do we feel about being replaced? Of course, this is not the first time that people have been made redundant in place of machines. Conveyor belts and very simple industrial robots and computers have already caused one level of redundancy. But still we have companies with human labor forces, particularly in China (e.g., FoxConn) that are being accused of employee exploitation such as 100 h of overtime per month for employees (3 times the acceptable rate) [item 62) in the Appendix]. Seemingly, we have never been so advanced as in this current stage of history, and yet we have never been so willing to bail out from the prospects of our own advancements? [item 63) in the Appendix] Is it because we are betting our dollar on the vision of a dystopic "machine-" driven world? Can we not see beyond? AI/AS has always meant to alleviate the human burden in processes. But somehow it seems to be propelling an even greater burden of digital laboring.

As these new emerging AS are packed with sensors and can almost see, hear, and touch through the recording of bits and bytes, we are left with the temporal that can now be stored and retained indefinitely, played back, audited, manipulated, and used in retrospective contexts [item 64) in the Appendix].

TABLE VIII
UNCHECKED AI/AS EXTERNALITIES

Unchecked Trajectories	Description
Conversational analysis	Data collected by IOT devices from speech to text
Bots mapping household boundaries	When personal mapping spaces are sent back via IOT
Cameras	When street cameras conduct visual analytics and are not just surveillance mechanisms
Billboards that study onlookers	Observers are having their privacy taken away and data used to further manipulate purchasers
Location sensors routing path	When location information is being onsold back to leaseholders of space in a shopping mall to tell shopkeepers who is browsing, at what time of day, and perhaps even what they are likely to purchase more of
Social media	Deep mining of sentiment analysis being conducted and then sold on to online retailers or governments
Ridesharing services	Location information is being collected and shared with third parties and then used to determine how much to charge someone on another trip

How do users know what is happening through blackbox technologies? Might the blockchain assist in these contexts and at what level of detail?

This is indeed the "fallout" from our intelligent products and processes with very deep social human implications [item 65) in the Appendix]. We can talk of these implications as unintended externalities that have the capacity to press against what in essence makes us humans (see Table VIII). Consider the conversation captured by Amazon's Alexa device, and accidentally sent to a random individual [item 66) in the Appendix]. Consider the iRobot Roomba device now mapping household physical and logical architectures [item 67) in the Appendix]. Consider the My Friend Cayla doll repeating swear words to a child that was hacked by a penetration tester [item 68) in the Appendix] or Hello Barbie's conversations being analyzed by Mattel's ToyTalk server? We could argue the fallout is limited because there was no direct "human" on the other side but what of domestic violence or cyberstalking victims who are continuously taunted by remote AI home-based attacks. What is frightening in the longer term is when we will not know what is actually driving decision making in AI/AS-based artifacts—might it be that sales are down that week? Or that upgrades might denote a new purchase? That back-end business alliances may be formed to onsell value-added products between a variety of retail chains? In short, this is the dark side of AI/AS and left unchecked will have dire consequences far beyond those reported in the Facebook-Cambridge Analytica scandal. This is why, we need global governance frameworks that have oversight over government, business, and consumer uses of AI/AS.

IX. GOVERNANCE AND ETHICAL FRAMEWORKS AND ACCOUNTABILITY SYSTEMS

If ethics exists at multiple layers of human engagement in the engineering process, and also at the system and subsystem level, how might we ensure that everyone and everything partakes and upholds their responsibilities? How do we ascertain that an AI/AS is truly compatible with the market it will be deployed in? Do we continue to rely on copious checklists of compliance by assurance personnel? Or will we create new AI/AS technology assessment tools identifying benefits and harms? Transparent communications, potentially through a blockchain register might well be how future complex systems are continuously assessed for adherence to fundamental principles devised by international and national bodies in harmony. There may also be a trend toward more open reproduceable and accountable data-driven processes that can be judged by anyone for their adherence to fundamental flows of information. The only risk of requiring systems to document systems-level transactions is the auditability and transparency of flows publicly available on the blockchain to communities of interest. There also need to be harsher penalties for those entities found in breach of shared resources. Smart energy suppliers in particular will have to prove to consumers that they will utilize streaming data from the home for positive benefits, and not to figure out what best new tariff mechanism to charge individual households to recoup the same amount of money once expended prior to smart solutions. Trust in suppliers and systems are among the most fundamental in our hope to make AI/AS central to our future [item 69) in the Appendix].

X. IN THIS ISSUE OF TTS

In this issue of TTS, we celebrate the inclusion of papers related to artificial intelligence and autonomous systems. The first paper is a collaboration between Anthony Aguirre of UCSC, Gaia Dempsey of the Seventh Future, Harry Surden of the University of Colorado, Boulder, and Peter Reiner of The University of British Columbia. Their paper titled: "AI

Loyalty: A New Paradigm for Aligning Stakeholder Interests," asks the fundamental question of whether we should assume that when we interact with an artificial intelligence system, that it is acting in our best interests, just like when we consult a doctor, lawyer, or financial advisor. The second paper is a single-authored contribution by Stamatis Karnouskos of SAP, titled: "Artificial Intelligence in Digital Media: The Era of Deepfakes." Following on from Aguirre *et al.*'s contribution, Karnouskos investigates deepfakes via multipronged perspectives that include media and society, law, and regulation. He underscores that as a society we are not ready to deal with deepfakes at any level, bringing to bear the importance of paper 1 in AI loyalty.

Paper 3 is by Tony Gillespie and Steven Hailes from University College London. They have contributed a paper focused on the assignment of legal responsibilities for decisions by autonomous cars using system architectures. Our final paper for this issue is another sole-authored paper and our first by a student member, Scott Greenhorn from KU Leuven, titled: "Toward a Secure Platform for Brain-Connected Devices—Issues and Current Solutions." Greenhorn, outlines some of the unique challenges that brain–computer interface-based devices will bring upon society. Its connection with this editorial has much to do with the future of IoT, the two types of black boxes described in this editorial. One can only imagine the possibility of security hacks in the brain, blocking the "human in the loop" capability, or riddling the AI in the brain with "deep fakes." In the future, could we distinguish between the loyal human and the loyal AI? We hope you will enjoy this special issue on AI/AS and look forward to your feedback.

ACKNOWLEDGMENT

The authors would like to acknowledge the copyediting of Dr. Rebecca Monteleone and Terri Bookman, and the critical eye of Associate Editor John C. Havens, and Publications Board Member Professor Philip Koopman.

KATINA MICHAEL
Arizona State University
Tempe, AZ 85281 USA

ROBA ABBAS
University of Wollongong
Wollongong, NSW 2522, Australia

GEORGE ROUSSOS
University of London
London WC1E 7HU, U.K.

EUSEBIO SCORNAVACCA
University of Baltimore
Baltimore, MD 21202 USA

SAMUEL FOSSO-WAMBA
Toulouse Business School
31068 Toulouse, France

APPENDIX
RELATED WORK

1) K. Michael and K. W. Miller, "Big data: New opportunities and new challenges," *Computer*, vol. 46, no. 6, pp. 22–24, Jun. 2013.
2) R. Pringle, K. Michael, and M. G. Michael, "Unintended consequences of living with AI," *IEEE Technol. Soc. Mag.*, vol. 35, no. 4, pp. 17–21, Dec. 2016.
3) M. G. Michael and K. Michael, "Resistance is not futile, nil desperandum [editorial]," *IEEE Technol. Soc. Mag.*, vol. 34, no. 3, pp. 10–13, Sep. 2015.
4) S. Hawking, *AI Could Spell End of The Human Race*, BBC News, London, U.K., Dec. 2014. [Online]. Available: https://www.bbc.com/news/av/science-environment-30289705/stephen-hawking-ai-could-spell-end-of-the-human-race
5) B. Kumpf, *Who is Writing The Future? Designing Infrastructure for Ethical AI*, Medium, San Francisco, CA, USA, Jun. 2018. [Online]. Available: https://medium.com/@UNDP/who-is-writing-the-future-designing-infrastructure-for-ethical-ai-4999620db295
6) S. Sackur, *Alan Winfield*, BBC World News HardTalk, London, U.K., Oct. 2017. [Online]. Available: https://www.bbc.co.uk/programmes/n3ct2km5
7) J. Pitt, ed., *This Pervasive Day: The Potential and Perils of Pervasive Computing*. London, U.K.: Imperial Coll. Press, 2012.
8) B. Ambrosino, *What Would it Mean for AI to Have A Soul?*, BBC Future, London. U.K., Jun. 2018. [Online]. Available: http://www.bbc.com/future/story/20180615-can-artificial-intelligence-have-a-soul-and-religion
9) M. Y. Vardi, "The moral imperative of artificial intelligence," *Commun. ACM*, vol. 59, no. 5, p. 5, May 2016. [Online]. Available: https://cacm.acm.org/magazines/2016/5/201608-the-moral-imperative-of-artificial-intelligence/fulltext
10) D. Lee, *IBM's Machine Argues, Pretty Convincingly, With Humans*, BBC Technol., London, U.K., Jun. 2018. [Online]. Available: https://www.bbc.com/news/technology-44531132
11) J. Pitt, ed., *The Computer After Me: Awareness and Self-Awareness in Autonomic Systems*. London, U.K.: Imperial Coll. London, 2014.
12) K. Ware, M. G. Michael, and K. Michael, "Religion, science, and technology: An interview with metropolitan Kallistos ware," *IEEE Technol. Soc. Mag.*, vol. 36, no. 1, pp. 22–26, Mar. 2017.
13) E. Guglielmelli, "Robots don't pray," *IEEE Technol. Soc. Mag.*, vol. 34, no. 3, pp. 15–16, Sep. 2015.
14) C. Wedge and C. Saldanha, *Robots Film*. Los Angeles, CA, USA: 20th Century Fox, 2005.
15) J. Badham, *WarGames Film*. Los Angeles, CA, USA: MGM/UA Entertainment, 1983.
16) K. Michael, "Human autonomy in emerging technology systems (ethics, policy, regulation track)," in *Proc. Comput. Community Catalyst Assured Auton. Workshop (CCC)* Phoenix, AZ, USA, Feb. 2020, pp. 1–38.
17) S. Akter, K. Michael, M. R. Uddin, G. McCarthy, and M. Rahman, "Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics," *Ann. Oper. Res.*, pp. 1–33, May 2020. [Online]. Available: https://doi.org/10.1007/s10479-020-03620-w
18) S. Bedaf, "The future is now: The potential of service robots in elderly care," PH.D. dissertation, Dept. Doctoral, Maastricht: Datawyse/Universitaire Pers Maastricht, Maastricht, The Netherlands, 2017. [Online]. Available: https://doi.org/10.26481/dis.20171221sb
19) T. Sorell and H. Draper, "Robot carers, ethics, and older people," *Ethics Inf. Technol.*, vol. 16, no. 3, pp. 183–195, Sep. 2014.
20) V. Evers, *The Rise of Social Robotics*, World Econ., Forum, Cologny, Switzerland, Feb. 2016. [Online]. Available: https://www.youtube.com/watch?v=KvgH-p6-Dt0
21) *Advanced Physical Security Technology Knightscope: K5*, Knightscope, Mountain View, CA, USA, 2019. [Online]. Available: http://knightscope.com/
22) K. Michael, *Meet Boston Dynamics' LS3—The Latest Robotic War Machine*, Convers., Parkville, VIC, Australia, Oct. 2012. [Online]. Available: https://theconversation.com/meet-boston-dynamics-ls3-the-latest-robotic-war-machine-9754
23) J. H. Moor, "The nature, importance, and difficulty of machine ethics," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 18–21, Jul./Aug. 2006.
24) R. Abbas, S. Marsh, and K. Milanovic, "Ethics and system design in a new era of human–computer interaction [guest editorial]," *IEEE Technol. Soc. Mag.*, vol. 38, no. 4, pp. 32–33, Dec. 2019, doi: 10.1109/MTS.2019.2948448.

25) A. Winfield, *Robotics: A Very Short Introduction*. Oxford, U.K.: Oxford Univ. Press, 2012.
26) A. F. Winfield, K. Michael, J. Pitt, and V. Evers, "Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]," *Proc. IEEE*, vol. 107, no. 3, pp. 509–517, Mar. 2019.
27) K. Michael, H. A. Love, and J. Wajcman, "Speaking out against socially destructive technologies: Norbert Wiener and the call for ethical engagement [guest editorial]," *IEEE Technol. Soc. Mag.*, vol. 36, no. 2, pp. 13–26, Jun. 2017.
28) R. Capurro and M. Nagenborg, *Ethics and Robotics*. Heidelberg, Germany: AKA G.M.B.H., 2009.
29) J. Borenstein, J. Herkert, and K. Miller, "Self-driving cars: Ethical responsibilities of design engineers," *IEEE Technol. Soc. Mag.*, vol. 36, no. 2, pp. 67–75, Jun. 2017.
30) C. Fisher and A. Lovell, *Business Ethics and Value S3rd Edition*. London, U.K.: Pearson, 2009.
31) *13482:2014 Robots and Robotic Devices—Safety Requirements for Personal Care Robots*, ISO, Geneva, Switzerland, Feb. 2014. [Online]. Available: https://www.iso.org/standard/53820.html
32) J. C. Havens and the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, Standard P7000, 2019. [Online]. Available: https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
33) *BS 8611:2016 Robots and Robotic Devices. Guide to The Ethical Design and Application of Robots and Robotic Systems*, AMT Committee Brit. Stand. Inst., London, U.K., 2016.
34) *Homepage*, Partnership on AI, San Francisco, CA, USA, 2018. [Online]. Available: https://www.partnershiponai.org
35) A. Koene, "Algorithmic bias: Addressing growing concerns [leading edge]," *IEEE Technol. Soc. Mag.*, vol. 36, no. 2, pp. 31–32, Jun. 2017.
36) K. W. Miller and D. Larson, "Agile software development: Human values and culture," *IEEE Technol. Soc. Mag.*, vol. 24, no. 4, pp. 36–42, Dec. 2005.
37) *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 1st ed., IEEE, Piscataway, NJ, USA, 2019. [Online]. Available: https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html
38) G. Marchant, *"Soft Law" Governance Of Artificial Intelligence*, AI Pulse, Tempe, AZ, USA, Jan. 2019. [Online]. Available: https://aipulse.org/soft-law-governance-of-artificial-intelligence/
39) D. Gotterbarn and K. W. Miller, "Unmasking your software's ethical risks," *IEEE Softw.*, vol. 27, no. 1, pp. 12–13, Jan./Feb. 2010.
40) L. Robertson, R. Abbas, G. Alici, A. Munoz, and K. Michael, "Engineering-based design methodology for embedding ethics in autonomous robots," *Proc. IEEE*, vol. 107, no. 3, pp. 582–599, Mar. 2019, doi: 10.1109/JPROC.2018.2889678.
41) A. Cavoukian, *Global Privacy and Security by Design*, GPS, Toronto, ON, Canada, 2019. [Online]. Available: https://gpsbydesign.org/
42) K. Michael, "Mental health, implantables, and side effects [editorial]," *IEEE Technol. Soc. Mag.*, vol. 34, no. 2, pp. 5–17, Jun. 2015.
43) J. Carvalko, *Death by Internet*. Mechanicsburg, PA, USA: Sunbury Press, 2016.
44) E. Strickland, *Layoffs at Watson Health Reveal IBM's Problem with AI*, IEEE Spect, New York, NY, USA, Jun. 2018. [Online]. Available: https://spectrum.ieee.org/the-human-os/robotics/artificial-intelligence/layoffs-at-watson-health-reveal-ibms-problem-with-ai
45) K. Michael and M. G. Michael, "The social and behavioural implications of location-based services," *J. Location Based Serv.*, vol. 5, nos. 3–4, pp. 121–137. Dec. 2011.
46) M. G. Michael, *Uberveillance: Definition* (Australia's National Dictionary), 5th ed. Macquarie Dictionary, S. Butler, Ed., Sydney Univ., Sydney, NSW, Australia, 2009, p. 1094. [Online]. Available: http://works.bepress.com/kmichael/178/
47) M. G. Michael and K. Michael, "Uberveillance: 24/7 x 365 people tracking and monitoring," in *Proc. 29th Int. Conf. Data Prot. Privacy Commissioners*, Montreal, QC, Canada, Sep. 2007, pp. 1–35. [Online]. Available: http://www.privacyconference2007.gc.ca/workbooks/pres_wrkshop1_01_michael_e.pdf
48) K. D. Stephan, K. Michael, M. G. Michael, L. Jacob, and E. P. Anesta, "Social implications of technology: The past, the present, and the future," *Proc. IEEE*, vol. 100, pp. 1752–1781, May 2012.
49) K. Michael *et al.*, "Planetary-scale RFID services in an age of uberveillance," *Proc. IEEE*, vol. 98, no. 9, pp. 1663–1671, Sep. 2010.
50) S. Rodotà and R. C. Rapporteurs. (Mar. 2005). *Ethical Aspects of ICT Implants in the Human Body: Opinion Presented to the Commission by the European Group on Ethics*. [Online]. Available: https://europa.eu/rapid/press-release_MEMO-05-97_en.htm?locale=en
51) K. W. Miller, "A secret sociotechnical system," *IT Prof.* vol. 15, no. 4, pp. 57–59, Jul./Aug. 2013.
52) M. G. Michael and K. Michael, "Toward a state of Überveillance [special section introduction]," *IEEE Technol. Soc. Mag.*, vol. 29, no. 2, pp. 9–16, Jun. 2010.
53) R. Abbas and K. Michael, "COVID-19 contact trace app deployments: Learnings from Australia and Singapore," *IEEE Consum. Electron. Mag.*, vol. 9, no. 5, pp. 65–70, Jun. 2020, doi: 10.1109/MCE.2020.3002490.
54) K. Michael "Brain pacemakers as consumer electronics for patient care: Benefits, risks and challenges (Part 1 interview with emeritus professor Gary Olhoeft)," *IEEE Consum. Electron. Mag.*, vol. 7, no. 4, pp. 82–85, Jul. 2018.
55) P. Hall, "Implantable technologies in the military sector," *IEEE Technol. Soc. Mag.*, vol. 36, no. 1, pp. 69–70, Mar. 2017.
56) K. Michael, M. G. Michael, J. C. Galliot, and R. Nicholls, "Socio-ethical implications of implantable technologies in the military sector [guest editorial]," *IEEE Technol. Soc. Mag.*, vol. 36, no. 1, pp. 5–8, Mar. 2017.
57) J. Ellul, *The Technological Society*. New York, NY, USA: Vintage Books, 1964.
58) G. Adamson, R. R. Kline, K. Michael, and M. G. Michael, "Wiener's cybernetics legacy and the growing need for the interdisciplinary approach [scanning our past]," *Proc. IEEE*, vol. 103, no. 11, pp. 2208–2214, Nov. 2015.
59) G. Wetware, *Profile*, Facebook, Menlo Park, CA, UA, 2012. [Online]. Available: https://www.facebook.com/GrindhouseWetware/
60) C. Metz, *Facebook's Race to Link Your Brain to a Computer Might be Unwinnable*, Wired, San Francisco, CA, USA, Apr. 2017. [Online]. Available: https://www.wired.com/2017/04/facebooks-race-link-brain-computer-might-unwinnable/
61) K. W. Miller, "Technology, unemployment, and power," *IT Prof.*, vol. 15, no. 6, pp. 10–11, Nov./Dec. 2013.
62) S. Liao, *Amazon and Foxconn Reportedly Strip Workers of Benefits and Pay*, Verge, New York, NY, USA, Jun. 2018. [Online]. Available: https://www.theverge.com/2018/6/11/17448544/amazon-foxconn-worker-conditions-benefits-stripped-low-wages-chinese-factory
63) M. Y. Vardi, *Are Robots Taking Our Jobs?*, Conversation, Parkville, VIC, Australia, Apr. 2016. [Online]. Available: https://theconversation.com/are-robots-taking-our-jobs-56537
64) J. Borenstein and K. Miller, "Robots and the Internet: Causes for concern," *IEEE Technol. Soc. Mag.*, vol. 32, no. 1, pp. 60–65, Mar. 2013.
65) M. G. Michael, K. Michael, "The fallout from emerging technologies: Surveillance, social networks, and suicide," *IEEE Technol. Soc. Mag.*, vol. 30, no. 3, pp. 13–17, Sep. 2011.
66) H. Shaban, *Amazon Echo Reportedly Recorded Conversation, Sent to Random Contact*, Sydney Morning Herald, Sydney, NSW, Australia, May 2018. [Online]. Available: https://www.smh.com.au/technology/amazon-echo-reportedly-recorded-conversation-sent-to-random-contact-20180525-p4zhev.html
67) M. Astor, *Your Roomba May Be Mapping Your Home, Collecting Data That Could be Shared*, New York Times, New York, NY, USA, Jul. 2017. [Online]. Available: https://www.nytimes.com/2017/07/25/technology/roomba-irobot-data-privacy.html
68) N. Oakley, *My Friend Cayla Doll Can be HACKED, Warns Expert—Watch Kids' toy Quote 50 Shades and Hannibal*, Mirror, London, U.K., Feb. 2015. [Online]. Available: https://www.mirror.co.uk/news/technology-science/technology/friend-cayla-doll-can-hacked-5110112
69) B. Shneiderman, "Design lessons from AI's two grand goals: Human emulation and useful applications," *IEEE Trans. Technol. Soc.*, vol. 1, no. 2, pp. 73–82, Jun. 2020, doi: 10.1109/TTS.2020.2992669.