2019

Engineering-Based Design Methodology for Embedding Ethics in Autonomous Robots

Lindsay Robertson
University of Wollongong, ljr894@uowmail.edu.au

Roba Abbas
University of Wollongong, roba@uow.edu.au

Gursel Alici
University of Wollongong, gursel@uow.edu.au

Albert Munoz
University of Wollongong, amunoz@uow.edu.au

Katina Michael
Arizona State University, katina@uow.edu.au

Publication Details

Robertson, L. J., Abbas, R., Alici, G., Munoz, A. & Michael, K. (2019). Engineering-Based Design Methodology for Embedding Ethics in Autonomous Robots. Proceedings of the IEEE, 107 (3), 582-599.

Engineering-Based Design Methodology for Embedding Ethics in Autonomous Robots

Abstract
This paper explores the design process of robotics and autonomous systems using a co-design approach, applied ethics, and values-driven methods. Specifically, the approach seeks to move beyond traditional risk assessment toward a greater consideration of end-user exposure. The goal of the ethics-based co-design approach is to identify end-user and stakeholder values that guide the minimization of end-user vulnerability associated with the employment of autonomous systems. This design process is also used to identify positive consequences that probably increase human wellbeing as opposed to simply avoiding harm. We argue that biomedical autonomous systems design, during the preclinical phase, should bring together diverse stakeholders that would not traditionally be involved in design. We also argue that embedding ethical considerations in the engineering design process should bring together a diverse range of stakeholders to more accurately appreciate possible end-user implications of a design. With complex systems design, such as biotechnologies, greater awareness is necessary of the ethical implications of designed autonomy to end-user exposure.

Disciplines
Business

Publication Details
Robertson, L. J., Abbas, R., Alici, G., Munoz, A. & Michael, K. (2019). Engineering-Based Design Methodology for Embedding Ethics in Autonomous Robots. Proceedings of the IEEE, 107 (3), 582-599.

INVITED PAPER

Engineering-Based Design Methodology for Embedding Ethics in Autonomous Robots

This paper explores a method for embedding ethics into the design and use of an endoscopic capsule for diagnosis and drug delivery, using a codesign approach to reduce end-user risk.

By Lindsay J. Robertson, Roba Abbas, Gursel Alici, Albert Munoz, and Katina Michael, *Senior Member IEEE*

ABSTRACT | This paper explores the design process of robotics and autonomous systems using a co-design approach, applied ethics, and values-driven methods. Specifically, the approach seeks to move beyond traditional risk assessment toward a greater consideration of end-user exposure. The goal of the ethics-based co-design approach is to identify end-user and stakeholder values that guide the minimization of end-user vulnerability associated with the employment of autonomous systems. This design process is also used to identify positive consequences that probably increase human wellbeing as opposed to simply avoiding harm. We argue that biomedical autonomous systems design, during the preclinical phase, should bring together diverse stakeholders that would not traditionally be involved in design. We also argue that embedding ethical considerations in the engineering design process should bring together a diverse range of stakeholders to more accurately appreciate possible end-user implications of a design. With complex systems design, such as biotechnologies, greater awareness is necessary of the ethical implications of designed autonomy to end-user exposure.

KEYWORDS | Autonomous robots; design; ethics; risk

I. INTRODUCTION

A. General Ethics Applied to Engineering Design—Nonautonomous

The design of machines, from an engineering perspective, is a relatively straightforward practice. Engineering design has required adherence to fundamental design principles, and an avoidance of unacceptable risk levels to ensure that the outcome will be functional, robust, and cost-effective. With increasing complexity and growing importance in society, machine (e.g., automobiles and airplanes) design necessitates end-user safety considerations to be deeply embedded in the engineering design process [1]. Automobiles were designed to include moving mechanical parts that require human interaction, and airplanes as they grew in sophistication over decades had some of the first computer systems with autopilot. Ethics, whether integrated into the engineering design process or not, has always been a topical area of inquiry for humanities and social sciences scholars [2]. Future visions of how autonomous systems might obviate human control [3] and replace fundamental human functions [4] have stirred debate regarding the interaction between humans and machines. This trend has meant that potential breaches in engineering ethics have received more attention than the potential methodologies that can embed ethics in the design of complex systems.

Ethics guide several important considerations in engineering design. First, ethics informs decisions about

Manuscript received March 28, 2018; revised August 21, 2018 and December 5, 2018; accepted December 20, 2018. Date of publication January 21, 2019; date of current version March 6, 2019. *(Corresponding author: Katina Michael.)*

L. J. Robertson is with the School of Computing and Information Technology, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: lindsay@tech-vantage.com).

R. Abbas and **A. Munoz** are with the School of Management, Operations and Marketing, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: roba@uow.edu.au; amunoz@uow.edu.au).

G. Alici is with the School of Mechanical, Materials, Mechatronic and Biomedical Engineering, University of Wollongong, Wollongong, NSW 2522, Australia, and also with the ARC Center of Excellence for Electromaterials Science, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: gursel@uow.edu.au).

K. Michael is with the School for the Future of Innovation in Society, Arizona State University, Tempe, AZ 85287 USA, and also with the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85287 USA, on leave from the School of Computing and Information Technology, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: katina.michael@asu.edu).

Digital Object Identifier 10.1109/JPROC.2018.2889678

whether to pursue research and development efforts to technological breakthroughs or continued reliance on the status quo. Second, ethics addresses important questions about what is being sought to build and why, the implications toward humans, and broadly guiding intent of the engineering efforts. Third, ethics has a role to play in the engineering design process, whether it is a functional feature in a single subsystem or is a feature that pervades the end-to-end system. Fourth, ethics allows the measurement of performance in the algorithmic source code of machines, as it relates to risks. Finally, ethics informs the dissemination of and access to particular technologies in terms of affordability [5].

Efforts to tackle how ethics can be applied to engineering design often include development of standards and an appreciation of potential tradeoffs. An example is the IEEE P7000 series of standards for "ethically aligned design" [6]. Also, country-specific standards are being developed to address machines such as robots in the context of ethical risk assessment. Examples include: adopting sociotechnical and ethical, legal, and social implications approaches, and other International Organization for Standardization (ISO) risk standards, such as ISO 31000. Although there appears to be considerable drive to develop guidance for engineering efforts in this space, there is little clarity about how ethics can be embedded into engineering design. On the one hand, machine related injuries or human fatalities tend to emphasize the need for ethical risk assessment and better complex systems design. On the other hand, profit-seeking industries strive to be the first to market, in order to capitalize on expected competitive advantage [7], [8]. These time pressures result in a tendency to adopt methodologies that embrace experimentalism.

As technology is increasingly embedded within the Internet of Things and humans, questions arise as to whether ethically aligned design processes are sufficient [78]. New patents are awarded for inventiveness, but do not require the submission of an ethics application [79]. And yet, it is the end-users who adopt machines who may be unnecessarily exposed to risks arising from the failure to apply engineering design ethics. Furthermore, utilizing metrics about end-user vulnerabilities may yield new ways of evaluating the ethics of updated models of machines. While we can never eliminate risk, we may be able to inform design toward the minimization of potential loss. This research posits that ethical considerations in the engineering design process should account for the vulnerability of specific end-users (e.g., patients and doctors). For a machine operating autonomously and making choices with ethical consequences, the design itself must be validated against a correspondent metric of vulnerability. Rather than rewarding inventiveness without considerations for the implications of added functionality, we argue that engineering design, and functionality improvements should be evaluated for their capability to minimize vulnerability when used.

B. Field of Application: Biomedical Engineering

This paper considers the design process for autonomous systems whose field of application has ethical implications: we consider the application of techniques that will ensure that the complete design process will result in confidence that ethical performance will be achieved. This paper does not consider detailed design, nor does it assess specific hazards. Rather, this paper considers the framework within which the design occurs, the stakeholder involvement in the design, the appreciation of limitations of autonomy, and the principles that will help to assure that the design will perform as expected when the operating environment is nonoptimal.

Throughout this paper, the authors will make use of the notion of a present mode of operation (PMO) and a future mode of operation (FMO) to distinguish between machine designs that are presently available as opposed to those that are a foreseeable future for the technology. It is likely, for a great number of devices on the market that a transition plan from a purely manual system to semiautonomous system and finally fully autonomous system has already been considered. The field of biomedical engineering is particularly prone to this type of technological trajectory given the precision required in instrumentation to be utilized. Examples currently exist in patient-controlled infusion pumps for drug delivery and by surgeons undertaking robot-assisted surgery (e.g., the minimally invasive da Vinci Surgical System).

A factor to consider is whether a device is a "brand new" technology (i.e., greenfield implementation) or a technology that is an incremental improvement on previous work (i.e., brownfield implementation). This distinction is significant as the process of development in biomedical applications must adhere to region-specific regulations where the innovation will be sold, many of which require acknowledgment of prior invention. In greenfield implementations, the initial design of a brand new device sets the course for further incremental innovation. If the right investment is made in a new technology from the initial design phase, then future generations of that technology may be similarly robust. Incremental innovations are less likely to achieve long term success if deployed at a lower threshold of ethical alignment.

If a technology is designed and built in a controlled environment, and tested only on nonhumans when its final application is meant to be in humans, then such a design will not go beyond a preclinical phase of development. Specific to the biomedical innovation, emphasis is placed on fundamental design controls with adequate historical documentation. As companies invest in technological transitioning from manual to autonomous systems, the design element will commensurately increase in complexity, especially in terms of embedded software engineering. Hybridized inventions that have dual functions (e.g., a medical diagnostic device that is capable of

Table 1 Alignment of Ethical Agency and Autonomy

Level of robot moral agency	Sheridan's autonomy	SAE Car autonomy
No moral agency	Computer offers no assistance; human does it all	Level Zero: No Automation You drive it.
No moral agency	Computer offers a complete set of action alternatives	Level One: Driver Assistance Hands on the wheel.
Implicit moral agent	Computer narrows the selection down to a few choices	
Implicit moral agent	Computer suggests a single action	
Implicit moral agent	Computer executes that action if human approves	Level Two: Partial Automation Hands off the wheel, eyes on the road.
Implicit moral agent	Computer allows the human limited time to veto before automatic execution	
Explicit moral agent	Computer executes automatically then necessarily informs the human	Level Three: Conditional Automation Hands off the wheel, eyes off the road - sometimes.
Fully moral agent	Computer informs human after automatic execution only if human asks	
Fully moral agent	Computer informs human after automatic execution only if it decides to	Level Four: High Automation Hands, off, eyes off, mind off - sometimes.
Fully moral agent	Computer decides everything and acts autonomously, ignoring the human	Level Five: Full Automation Steering wheel is optional.

drug delivery) are assessed by multiple panels within a regulatory body. Regulators in turn will need to seek new skilled expertise in this emerging area to assess biomedical designs that are potentially hazardous to humans if they do not meet respective standards. In the context of brownfield implementations of biomedical devices, such as implantable cardioverter defibrillators, incremental innovations may require designers to improve upon the incumbent architecture available toward vulnerability reduction, even if the adopted design comes from a competitor innovation. Furthermore, as complex semiautonomous or fully autonomous designs enter the market, a greater emphasis will be placed on the interconnection of these technologies with open infrastructures, such as the Internet and Internet of Things. Aspects of privacy and security that were not a major consideration in once-standalone devices such as heart and brain pacemakers will now receive greater attention [80].

C. Distinctive of Ethical Design Applied to (Semi)Autonomous Systems

Humans learn and absorb ethical values from parents, culture, and society (e.g., legal codes). Humans possess empathy and understand that harm to others has a negative impact. Humans also have self-worth, and hence the prospect of financial penalty, jail time or worse is a powerful incentive to conforming behavior. In addition, humans have autonomy, and as long as our tools have no autonomy we will exert specific ethical maxims and aim to avoid harm to ourselves. When humans consider the possibility of a robot acquiring some level of autonomy, we must question whether the robot will be able to perceive ethical issues, and incorporate mechanisms to ensure ethical behavior [9]. The human incentives for ethical behavior seem inapplicable—a threat of incarceration is unlikely to influence a robot's proposed actions—so we must also assess whether the capability of a robot to recognize ethical implications is well-aligned with its devolved level of autonomy.

Categorizations of autonomy have been proposed, including Sheridan's [10], and the categories of the self-driving car [3]. These categorizations could be applied quite generally, as can ethical agency [11]. While precise alignment is difficult, Table 1 (adapted from [3], [10] and [11]), attempts to show that it is possible to align the categorizations of ethical agency with categorizations of

autonomy. Thus, alignment gives essential guidance to the ethical design process by specifying acceptable levels of ethical agency according to the level of autonomy allowed in the design.

We propose that the design process must ensure that a given robot's level of autonomy to make decisions with ethical implications is adequately aligned with its capability to recognize situational actions with ethical implications [12]. For many design issues involving no devolution of autonomy, there is a possibility that an ethical human can intervene to mitigate or avoid harm.

There is a clear need for guidance that can be translated into engineering terms, to inform metric selection during the design process and to avoid inventing machines that increase the risk of ethical violations in an increasingly machine-autonomous world. This paper proposes an ethically aligned co-design methodology to provide such guidance, as detailed below and validated through the biomedical device case study.

This paper does not attempt to provide a "cookie cutter" approach to methodology. We address the major methodological functions by considering: 1) philosophically how to obtain agreement on functional requirements (co-design); 2) how engineering design could achieve the identified functionality; 3) how to clarify the applicable ethical issues; 4) how to ensure that autonomy levels proposed in the design are well-aligned with the level of ethical decision-making capability; and 5) a proposal of an approach to providing greater assurance that the final design will actually provide those capabilities even in adverse circumstances.

D. Outline

This paper explores the process of robotics and autonomous systems development using a co-design approach to reduce end-user risk. The theory of exposure is introduced to evaluate end-user vulnerability while undergoing endoscopic procedures using a robotic capsule with autonomous drug delivery capability. Underpinning this approach is an end-to-end ethically aligned co-design methodology engaging end-users throughout the design process aimed at hazard adjustment. Normally, endoscopy practice requires the use of invasive instrumentation to gather information necessary for diagnoses and treatment. The procedure can alternatively be conducted with a noninvasive ingestible capsule-shaped device containing a camera, LED lights, battery, and electronic circuits. The capsule allows real-time transmission of data to facilitate medical diagnosis while moving in the human gastrointestinal tract under natural peristalsis. Future designs will include active navigation for diagnosis, treatment, sample collection and site-specific drug delivery. The contribution of this paper is a method for embedding ethics into the design of a machine in a sociotechnical application, showing applicability to this rapidly developing field.

II. ETHICAL CO-DESIGN: A PHILOSOPHICAL APPROACH

A. Overview and History of Co-Design

Co-design refers to any system or technology design effort that engages end-users and other relevant stakeholders in the creative process, in an "act of collective creativity" [13]. The term "process" is important, as co-design is not to be misconstrued as a single or discrete "event" but rather an iterative and integrated practice that was originally aimed at end-users but now encompasses other stakeholders that may be affected by the implementation of a specific system or service [14].

The co-design approach has been considered and deployed in a variety of contexts, including: in the design of social services and public policy by including citizen participation [15] and [16], the early childhood education sector to inform the design of technology intended for children [17]–[24], in participatory design projects focused on collaborative urban planning [25], and in the biomedical (augmented reality and robotics) application user centric development [26] among others.

Historically, the co-design philosophy can be traced back to the European participatory design movement of the 1970s [27]. This movement elevated the importance of participation, cooperation and co-creation, generally within an open systems design [25], [28]. For a detailed overview and further readings on participatory design, fundamental to which are the ideas of participation, cooperation and co-creation, generally within an open systems paradigm (see [27]). Another notable concept is the distinction among a different "arenas for participation" [29], [30], specifically the individual, company and national (and even international) arenas, with certain challenges emerging when design activities enter the public sphere [31], [32]. Participatory design, although crucial in the early stages of the design process, faces some implementation issues due to constraints inherent in the real-world (e.g., time and budget). These constraints are also experienced in engineering design, as a tendency exists for isolated design; and the analysis of early prototypes that may not succeed are closed to end-users. This is particularly true if the creation and refinement of more mature prototypes are necessary prior to soliciting feedback.

In cases where user engagement is sought, user-centered design approaches are often preferred over partnerships, the latter being a basic tenet of the co-design concept. The co-design philosophy is centered on recognizing end-users as experts and partners as opposed to a conventional client-designer relationship [27], and user-centered approaches that value the "expert perspective" which considers the "user as subject" [13]. This is a fundamental distinction to make when comparing co-design with user-centered approaches, as the emphasis is on empowering users to become "experts in the experience domain" [33] and, therefore, being encouraged to

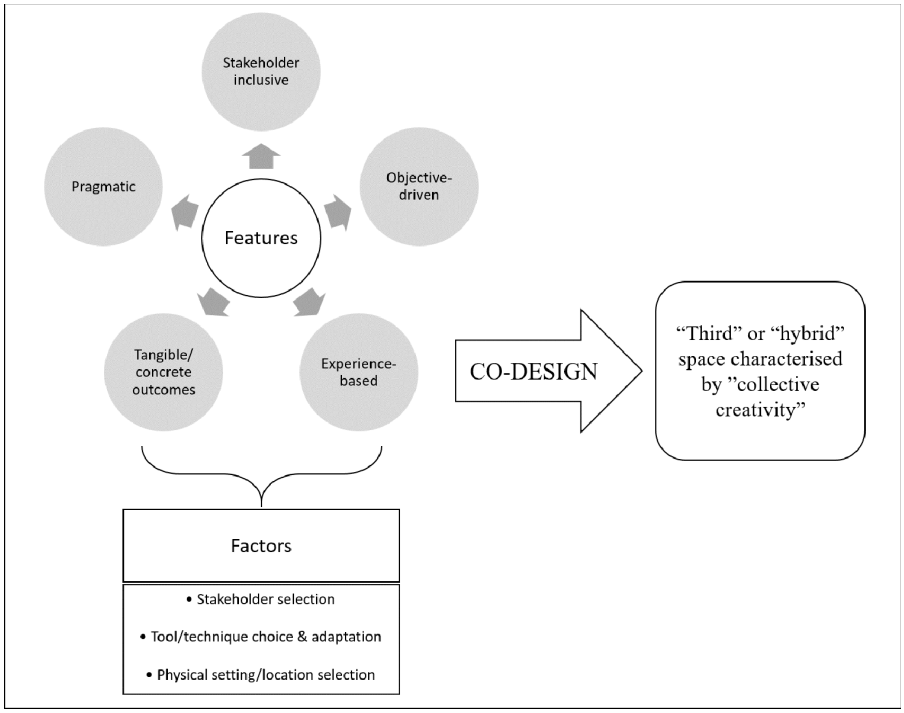

Fig. 1. ***Features and factors of co-design facilitating a shared creative experience.***

contribute as valid and important members of the design team.

B. Features and Factors of Co-Design

A successful co-design implementation must take into account three factors: 1) stakeholder selection, to ensure an appropriate participant set is chosen; 2) tool and technique choice, whereby existing co-design methods are adapted to the given project; and 3) physical setting selection to ensure the use of the most suitable location in which co-design activities can occur (adapted from [32]). Co-design must also focus on experiences as the basis for the design process in an inclusive, objective-driven, concrete, and pragmatic fashion adapted from [15] and [16] to create a "third space" (see [28] in which "collective creativity" [13] takes place and thrives). This third or "hybrid" space is "concerned with creating regions of overlap where the perspectives can come into mutual knowledge and, potentially, alliance—with the creation of the hybrid spaces in which objectivity can emerge through constructive discussion, dialog, negotiation, and mutual learning" [28]. A conceptual representation of the major features and factors of co-design guide a shared creative experience in a defined "third space" as shown in Fig. 1.

C. Applicability of Co-Design to Robotic Capsule Endoscopy: How to Achieve Ethical Alignment

To argue the applicability of co-design in the context of a robotic endoscopy capsule, and the development of an ethically aligned co-design methodology, the conceptual representation outlined must be extended to recognize aspects relevant to the (autonomous) robotic capsule. The socio-technical context in which robotic capsule endoscopy exists, falls well within the socio-technical theory realm. In socio-technical theory, interactions between humans and technology arise as primary drivers of system behavior. By extension, the co-design process must be acknowledged as being an open system, as performance outcomes influence, and are influenced by the end-user.

The co-design philosophy is often linked to the open systems paradigm, which in turn is associated with notions from socio-technical systems theory. Open systems, in their basic form, can be defined as systems that are receptive to information generated and provided by the environmental or external context [34]–[36]. The socio-technical approach similarly recognizes the importance of exogenous factors that influence a given socio-technical system. While much has been written about the socio-technical approach to design (refer to [37]–[41] for influential texts), elementary applications will pay particular attention to the social and technical elements that comprise a given system. These social and technical subsystems interact in an often inseparable manner to produce system outcomes. The socio-technical approach originated in the context of primary work systems in an organizational setting, but was later extended to encompass societal-level systems and an environmental subsystem (e.g., soft-regulations) that affect and influence socio-technical system outcomes.

An understanding of socio-technical principles is essential to any co-design effort, and the fundamental principle of interest is that of joint optimization, which stipulates that: the nature of interactions between the social and technical elements defines the degree of success that the socio-technical system as an entity achieves [42]. While traditionally, joint optimization efforts were concerned with the social and technical subsystems, the definition was extended in recognition of the environmental subsystems or factors [43].

D. Engineering Design Process

Engineering design and development processes must be acknowledged as part of this discussion, given that they drive the creation of the robotic capsule studied in this paper. Our insight into the engineering design process of the capsule is informed by the first-hand experience of the authors, and semiformal interviews of field experts. Previous studies have documented biomedical technology engineering design methodologies as following a linear sequential approach (e.g., waterfall model as per [44], [45]). An alternative co-design process based on user-centered design principles has also been applied [26]. However, ethical alignment was not a prominent feature of the suggested methodology and the relevance of existing processes and their role in the co-design method were not detailed. For the purpose of this paper, a representative engineering design process would include the following stages:

1) needs identification;
2) literature/background study;
3) task requirements and specifications;
4) definition of the goal/purpose of the design;
5) ideation and invention;
6) analysis;
7) selection;
8) detailed design;
9) prototyping and testing (including validation, certification and standardization as applicable);
10) production.

Design of biomedical machines must acknowledge progress in the ethics domain made by academia, regulatory and standards bodies. Specific to this area, extant literature provides clear definitions, regulatory frameworks and endeavors to minimize breaches through standardization of design procedures. Efforts to counter the possibility of adverse events are principally informed by evaluations of design feasibility [46]. In the case of biomedical technologies, these evaluations account for the potential for harm to the end-user (i.e., patient). Exposure broadly refers to the possible loci for hazards, and associated consequences. In involving end-users in the design process, a different perspective can be appreciated, considered and implemented into the design, particularly in the preclinical phase of development (i.e., prior to implementing the design process noted above in its entirety). In many cases, these implementations can dramatically reduce the possibility of end-user harm. These latter two considerations—biomedical machine ethics and end-user vulnerability—involve a considerable body of the literature and will be further explored as follows.

E. Machine Ethics: Scope

The co-design process must be ethically aligned and adhere to appropriate regulations and standards. Therefore, it is valuable to define the scope and background as to what it means to be ethically aligned in the context of (autonomous) machines in general, and the robotic capsule endoscopy systems more specifically.

In the Western world, controversial ethical issues (e.g., euthanasia and the infallibility of a Ruler) are commonly decided by a complex and interlinked system of judicial case-law and community principles. Decisions may be guided by legislation generated in response to community assessments of ethical issues, or by current community standards. This codification is somewhat circular, since community standards are codified into statutes, which in turn strongly influence community standards.

Alternatively, ethics can originate from compliance with an accepted moral imperative, ubiquitous among communities. One "moral law," formally presented by Kant [47] as the categorical imperative, specifies what individuals ought to or are obliged to do. This single moral rule stipulates that an individual must "act only on that maxim by which you can at the same time will that it should become a universal law." This has often been considered aligned with the "golden rule," which is thought to have originated in Confucian times (see Confucius in [48]), and is adopted as a fundamental directive by many religions and societies. The golden rule essentially states that you should treat others as you wish to be treated; a commonly held (but not the only) ethical position. Kant [47] also presented the second formulation of the categorical imperative, which recommends the following: "act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end." These issues come into focus when a robot is considered. Regardless of the level of image processing, deduction and memory incorporated into a robot, it is hard to envisage a situation where it would be acceptable to sacrifice the life of a child in order to protect the existence of some advanced example of consumer electronics. This is perhaps the quintessential difference between "robot" and "human." For example, NASA recently commanded its Cassini spacecraft to "commit suicide" by diving into Saturn, in order to prevent possible contamination of Titan with earth-originating bacteria. One could speculate on whether a more autonomous version of Cassini might have considered that it had a personal chance of survival by landing on Titan and was not too concerned by an academic consideration of human bugs.

Moor [11] considers "implicit ethical agents," "explicit ethical agents," and "full ethical agents."

Moor proposes: 1) that an "agent" whose scope of action is specifically designed responses to constrained issues that have ethical implication is an "implicit ethical agent"; 2) that an agent that can make limited autonomous ethical decisions is an "explicit ethical agent"; and 3) that an agent which can determine and justify ethical principles can be designated as a "full ethical agent." Moor also notes that "clear examples of machines acting as explicit ethical agents are elusive," and furthermore "an average adult human is a full ethical agent" [49] and does not nominate categories but concludes that a robot becomes a moral agent first, when "the robot is significantly autonomous from any programmers or operators of the machine." The second is when one can explain an autonomous machine's ethical violation only by ascribing it to malicious intent. And finally, moral agency requires "the machine to behave in a way that shows an understanding of responsibility to some other moral agent."

It is reasonable to consider an average human as a full ethical agent. This paper ascribes no particular religious bias to the "parable of the good Samaritan" (Luke 10:25–37) but a designer of an autonomous system might make note that even groups with identical cultural backgrounds and an unreservedly common adherence to a form of words could apparently arrive at significantly different interpretations of the appropriate actions.

The Autonomy Levels for Unmanned Systems (ALFUS) framework [50] has proposed a finer-grained definition of autonomy, noting three dimensions: 1) mission complexity; 2) environmental complexity; and 3) human independence or autonomy level. If the design process considers the ALFUS levels of autonomy against ethical criteria, it may conclude that there is a need to reduce the ALFUS autonomy levels permissible within a design to the point where the ethical requirements can be met. Thus, such ethical criteria would be necessary before a human designer could reasonable devolve to a robot, more than an "implicit agent" level of autonomy [11].

In order to examine the applicability of ethical decisions in the design of robots, we can perhaps separate some functional elements.

1) Can a human designer codify conditions under which a specified action will benefit a human?
2) Can a human designer codify relevant issues which a representative person would perceive as harmful (physical harm, privacy, humiliation, and embarrassment), and quantify/categorize these to a degree that could allow decision-making?
3) Can a human designer codify relevant environmental conditions that will modify perceived levels of harm? If I am shipwrecked naked, I would rank the harm of nonrescue as greater than the harm of appearing naked?
4) Can a human designer quantify relevant situations where a robot action will cause differing types and degrees of harm to more than one person? That is, minor embarrassment to one person, lethal danger to a large number or long-term high-risk to others?
5) Can the robot identify/quantify all relevant human harms and harm-levels? Does the robot have, within its construction and computational abilities, the capacity to identify all relevant types of harm?
6) Can the target robot predict, from alternative actions, the levels and types of harmful effects that those actions will cause for each potentially affected person?
7) Does the human designer who constructs/programs the robot have the capability to imbue these recognition capabilities?
8) Is the robot capable of autonomously choosing to carry out actions that could potentially cause various "harms" to one or more persons?
9) Is the robot capable of examining choices available to it, including choices to terminate its own existence? And to determine levels of identifiable harms likely to arise for each of the full range of potentially affected persons, from each alternative robotic choice?
10) If a robot is able to select from a range of actions (including a selection of inaction that could potentially cause varying levels of types of harm to one or more persons), are the definitions of harm and the numbers of persons and the environmental modifying issues sufficiently quantifiable to allow decisions that would be acceptable to society?
11) Can the robot apply these principles statistically, i.e., taking the view that it will make "correct" decisions 90% of the time, and for 10% of the time its decisions will prove to be incorrect and harmful?"

None of these evaluations are trivial: Dennis *et al.* [12] propose a calculus that would include a "safety/ethical logic layer (SEL)." The authors constructed a model capable of predicting the consequences of a robot's actions [51], embodying acceptable ethical principles. Specifically, the robot proactively selects actions that will minimize risk providing a basis for a declarative language that selects actions based on the evaluation of ethical consequences [12]. This is undoubtedly a valuable approach, addressing point 6 above.

In those authors' example, the harm was very simple: keeping a person from falling down a hole (refer also to [52]). Within a robot of the type envisaged for this paper's case study, these principles are also relatively simple to apply, but in more complex and common cases the application of Dennis *et al.*'s [12] approach needs to be tested.

The designer can then consider the ALFUS levels of autonomy, against the above criteria, and conclude that there is a need to reduce the ALFUS autonomy levels permissible within a design, to the point where the ethical requirements are able to be satisfied. As such, specific criteria, generally impossible to fulfill at present, would be necessary, before a human designer could reasonably

devolve to a robot, more than an "implicit agent" [11] level of autonomy.

F. Standards Applications to Autonomous Systems

A number of international standards have already emerged, and professional bodies such as IEEE have instigated working groups (notably the "Ethically Aligned Design" [Versions 1 and 2] documents arising from the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems) to collect and clarify the numerous relevant publications and approaches available. In the field of autonomous systems, IEEE P7000 "The First Global Standard Process for Addressing Ethical Concerns in System Design" offers "a methodology for identifying, analyzing, and reconciling ethical concerns of end users at the beginning of systems and software life cycles. The purpose of IEEE P7000 is to enable the pragmatic application of this type of value-based system design methodology which demonstrates that conceptual analysis of values and an extensive feasibility analysis can help to refine ethical system requirements in systems and software life cycles..." Central to the theme of this paper is BS 8611:2016 "robots and robotic devices. Guide to the ethical design and application of robots and robotic systems."

Vagia *et al.* [53] present a literature review on the levels of automation identified over the years. Sheridan and Verplank [10] proposed ten gradations of autonomy that are widely used: these ranges from "Computer offers no assistance; human does it all" to "Computer decides everything and acts autonomously, ignoring the human." The "ALFUS" [50] ("NIST/ALFUS") are commonly used and provide independently assessed rankings under three categories: human independence, mission complexity, and environmental complexity. Initially, the SAE [3] nominated six levels of autonomy for cars, ranging from "level 0: no automation. You drive it. ... Level 5: full automation. Steering wheel is optional. The front seats might face backwards." However more recent publications, for example, the SAE International On-Road Automated Vehicle Standards Committee [3] Information Report: (J3016) "Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems," focuses on higher levels of automation.

G. Vulnerability, Exposure, and Risk

A number of metrics and the respective procedures exist in the risk analysis literature. From a theoretical perspective, evaluations about risk can be embarked upon in a number of ways, many of which are far more complex than the multiplication of event occurrence probability by the consequences of the event, summed across a set of possible scenarios [54]. Ultimately, the method of evaluation of risk, as in any measurement exercise, should correspond to the aims and objectives of the study while being adequate and acceptable for decision making [54], [55].

Specific to the operation of technology in socio-technical systems, functional safety standards exist to guide practitioners in ensuring a piece of technology operates within acceptable levels of risk (e.g., IEC EN 61508). Generally, such standards are developed to provide a generic method of identification, classification, mitigation of hazards, and guide contingency protocols (e.g., safety case). Embedded within these standards are a number of risk assessment methods that aim to either amalgamate two or more dimensions of risk into a single index measure of risk (e.g., risk matrices and safety integrity levels [56]), or specify the consequences of a given event (e.g., fault tree analysis and failure mode and effects analysis). Certainly, many of these standards are applicable and valuable to the mitigation of any vulnerability generated in the design process. However, if the scope of the study extends beyond the extent of the standards, then a more tailored approach may be required (see [57]). For example, a design that employs functional safety practice may adopt the use of risk matrices to categorize likelihood and consequences of failures of a biomedical device to deliver an output by amalgamating both variables into a measure of risk given a type of failure. This approach may not be fit for purpose given the context, as the relative importance of likelihoods and consequences will differ considerably in medical settings (see [58]–[60] for a detailed discussion).

In practice, a number of international and national standards are applicable to the analysis of risk: AS/NZS ISO 31000:2009, MIL-STD-882E, DO-178C:2011, IEC 61508:2010, DEF-STAN 00-56 Issue 4 (2007), ISO/WD PAS 21448, among others. In addition, the Risk Management: Principles and guidelines, and AS/NZS ISO 31000:2009. Noted, the ISO GUIDE 73:2009 Risk management: Vocabulary have replaced AS/NZS 4360:2004 and represent the industry standards for risk-related terminology. IEC 61508:Edn 2 (2010) framework and associated IEC 61511, IEC 61513, and IEC 62061 standards for engineering practice related to functional safety of engineering systems. There is also the ISO/IEC 26702:2007 (replacing IEEE Std 1220-1998) Systems engineering: Application and management of the systems engineering process which is relevant to the design aspects of a broad range of systems (and certainly the endoscopic capsule system presented in this paper). ISO 45001:2018 Occupational health and safety (replacing BS OHSAS 18001:2007) is relevant to the working conditions of the health care service provider (e.g., surgeon). For a case where human injury is possible, failure mode effect and criticality analysis are also relevant to the design process, although the United States Department of Defense (1980) MIL-STD-1629A is no longer in force. ISO 13485:2016 specifies "requirements for a quality management system where an organization needs to demonstrate its ability to provide medical devices and related services that consistently meet customer and applicable regulatory requirements. Such organizations can be involved in one or more stages of the life-cycle,

including design and development," and ISO 14971:2007 provides specific advice regarding the application of risk management processes to medical devices.

Within a socio-technical system where humans interact with partially automated technologies, an end-user is vulnerable to failures of both the human and the technology. Thus, end-user exposure to output failure is contingent upon both the human and automated technology failing. Conversely, end-users engaging with an autonomous system subject to analogous failures will have a comparatively higher exposure level as a single failure will cause failure to appropriately deliver the expected output, lacking the option for human intervention. Although intuitively obvious, an exposure analysis is valuable if included in the design process. An exposure analysis that employs a metric of end-user exposure capable of attributing variations across measurements to specific contributors can aid the development of designs with reduced end-user vulnerability.

H. Exposure

The analysis of exposure is presented in this paper as a core feature of the co-design method, and it is useful to define the term and the manner in which it will contribute to the proposed, ethically aligned co-design methodology. Poorly designed or integrated components of a technological system represent weaknesses that can result in threats to the system capability to deliver goods or services as designed. Exposure, therefore, represents an evaluation of the contact potential between a hazard and a receptor [61]. A threat to an end-user, engaging with a technological system is only significant if it aligns with a specific weakness of that system resulting in contact that leads to exposure. Conversely, every weakness can potentially be targeted by a threat—either external or arising from a component's failure to achieve "fitness for purpose"—and so the configuration of the system's weaknesses influences the end-user's "exposure." This suggests the importance of system configuration in determining the vulnerability of end-users, and the importance of an approach based on a transparent and auditable mapping to the actual system. If a service level output from a system is defined, it is possible to describe the complete system's output using a Boolean expression, with AND functions describing cases where an intermediate stream is generated when input streams and functional processes are available, and OR functions describing cases where multiple processes or streams can supply the required functionality. A truth table representation of inputs and output allows cases where single failures cause output failure to be summed, and similarly allows a summation of the number of cases where dual failures cause output failure, etc. These summations can be represented as $\{E_1, E_2, E_3 \ldots\}$ where E_1 is the number of cases where a single stream or process failure will cause output failure, and E_2 is the number of cases where two failures (neither contributing to E_1) will cause output failure etc. For a potentially life-threatening output from an autonomous machine, it would seem reasonable to stipulate that $E_1 < 1$ and $E_2 < 1$. Given the context, it is reasonable to assume that most if not all output failures have sufficiently severe consequences to merit consideration during design decisions to avoid output failure possibility. The metric $\{E_1, E_2, E_3 \ldots\}$ has previously [59] been shown to be a representation of the "exposure" of the technological system from which it is derived, and a valid measure of the vulnerability imposed on the user. Perrow [62] asserted that whenever a system is sufficiently complex, failure is inevitable; however, the evaluation of the system under consideration offers opportunities for both identifying and assessing potential improvements.

Any "design" is at least implicitly validated only for a particular operating environment or context (e.g., NASA's "curiosity" Mars rover would not operate on the surface of Venus). The co-design process described in this paper can be expected to result in a design that operates under normal circumstances, and is very likely to perhaps quite unconsciously incorporate the designers' perceptions of risk. Previous research [63] has noted specific limitations inherent in the concept of risk, and for an autonomous system that has the potential to make ethically significant decisions, a risk-based approach to design is proposed to be inadequate for two reasons: first, the difficulty of ensuring that all hazards are identified, second, the potential for failure to identify specific weaknesses, and third, the difficulty of justifying risk probabilities based on expert assessment.

Rather than simply identifying individual possibilities of failures, this approach considers all component/systems and the combinatorial (as well as individual) possibilities of failure created by the actual design. Mapping the configuration of components and subsystems also facilitates review and validation, and is relevant to both the PMO and FMO.

An analysis of the "exposure" of the system provides a numerical and defensible measure of the weaknesses of the autonomous system, and contributes directly to the design process. If there are known weaknesses in the architectural and system design of the product/process, then address these before diffusion into the market. As one example of a field of contribution, it can be observed that if a final design goal were to be elimination of E_2 vulnerabilities, then providing design redundancy on both of the two components that contribute to an E_2 vulnerability would be wasted as the same effect could be achieved by providing design redundancy to either one of the two vulnerabilities. Based on projects other than that illustrated in this paper, it is common to find that the total exposure of a system can be dramatically reduced by eliminating a design requirement for a highly exposed contributory system (e.g., a mandatory internet connection). The numerical analysis of a system's exposure requires a pragmatic approach to the granularity of design that is used; Robertson [58] proposes practical criteria for granularity to be considered, and also

requires that "fail" is defined in terms of nondelivery of outputs required by a following or final usage.

The measure of exposure associated with a design, can then serve as a feedback mechanism to evaluate weaknesses in a design, and ensure a final design meets an agreed level of safety. The measure of exposure allows that hazards are appreciated and reduced to a level no higher than could be offered by an equivalent, less autonomous approach. The analysis of exposure provides an essential step in the design process that forces the consideration of weaknesses rather than the designer's perception of risk. A socio-technical system design claimed to be ethical, needs to demonstrate objectively that the designer rigorously considered and evaluated design issues and possibilities. Specific to the case, risk consequences should be considered regardless of probability, rather than estimating probability of occurrence and consequences in aggregate indices, and thus reproducing the issues highlighted by Cox [60].

III. PROPOSED CO-DESIGN METHODOLOGY

A. Fundamental Principles and Concepts

The theoretical insights gained from the literature review coupled with practical experience acquired by engineers and technologists consulted that are currently working in this domain, have supported the emergence of a series of fundamental principles and concepts that must be integrated into the proposed ethically aligned co-design methodology. These principles and concepts include in the following.

1) The (engineering-based) system is an open system, in a theoretical sense, whereby interactions occur in a broader socio-technical context. Environmental factors exert a direct influence on the system, through the provision and exchange of information.
2) The socio-technical system in question is largely influenced by existing engineering design processes, which are often in progress when a co-design methodology of this nature is put into practice. Therefore, the appreciation and integration of existing engineering design frameworks is critical.
3) Engineering design processes operate within a wider innovation setting, characterized by three distinct but interrelated phases; prior to and during the introduction to the market, and after the innovation becomes available to end-users.
4) The socio-technical system, as made up of inextricably linked social and technical subsystems within a unique environmental context, must be considered at various levels throughout the design process.
5) Relevant stakeholders, notably end-users, should be actively involved during the engineering design process, and at each of the aforementioned levels of design.
6) Stakeholder engagement should not be restricted to end-user involvement, but should encourage and support the inclusion of additional stakeholder groups who may be influenced by the engineering design.
7) For the co-design process to be ethically aligned, a thorough understanding of the existing regulatory environment is required to facilitate integration of such provisions in the early stages of the engineering co-design process.
8) A standard risk assessment has inherent limitations that are particularly relevant to this application. Rather, underlying the co-design methodology is the analysis of "exposure" as a metric of system weaknesses that serves as feedback during the design process, through the provision of contextually relevant measurements that embody risk in use.

B. Application of the Methodology to Biomedical Engineering and Robotic Capsule Endoscopy

The application of the aforementioned principles and concepts to the biomedical engineering field, and specifically to robotic capsule endoscopy, requires a number of assumptions be made with respect to the case study presented and examined for the remainder of this paper.

1) A robotic capsule system comprising the robotic capsule (device) and the social, technical, and environmental contexts in which it exists, is an open unit that is directly influenced by, and is receptive to, changes in its surroundings. It does not, and should not, exist or be designed and developed in isolation.
2) The creation of a robotic capsule requires awareness of typical engineering design (and to some degree, development) processes. Preliminary stages of such processes include some form of needs identification, background and literature study, requirements specification, the identification of the objectives of the design, and an ideation component. These preliminary phases are followed by prototyping with a focus on exhaustive analysis of multiple designs. Such analysis in turn informs the selection of a preferred prototype leading to a detailed design phase. The latter is concerned with the construction and exhaustive testing of the selected prototype, culminating in the production phase of engineering design.
3) The innovation setting for the co-design of the robotic capsule is comprised of the preclinical, clinical, and diffusion phases. In reality, the analysis, selection and detailed design of prototypes form the crux of the preclinical phase of development and is where the proposed co-design methodology should be introduced. Ideally, however, co-design activities would span the entirety of both the engineering design and innovation processes.
4) The co-design of an (autonomous) robotic capsule should be considered at various levels, ranging from

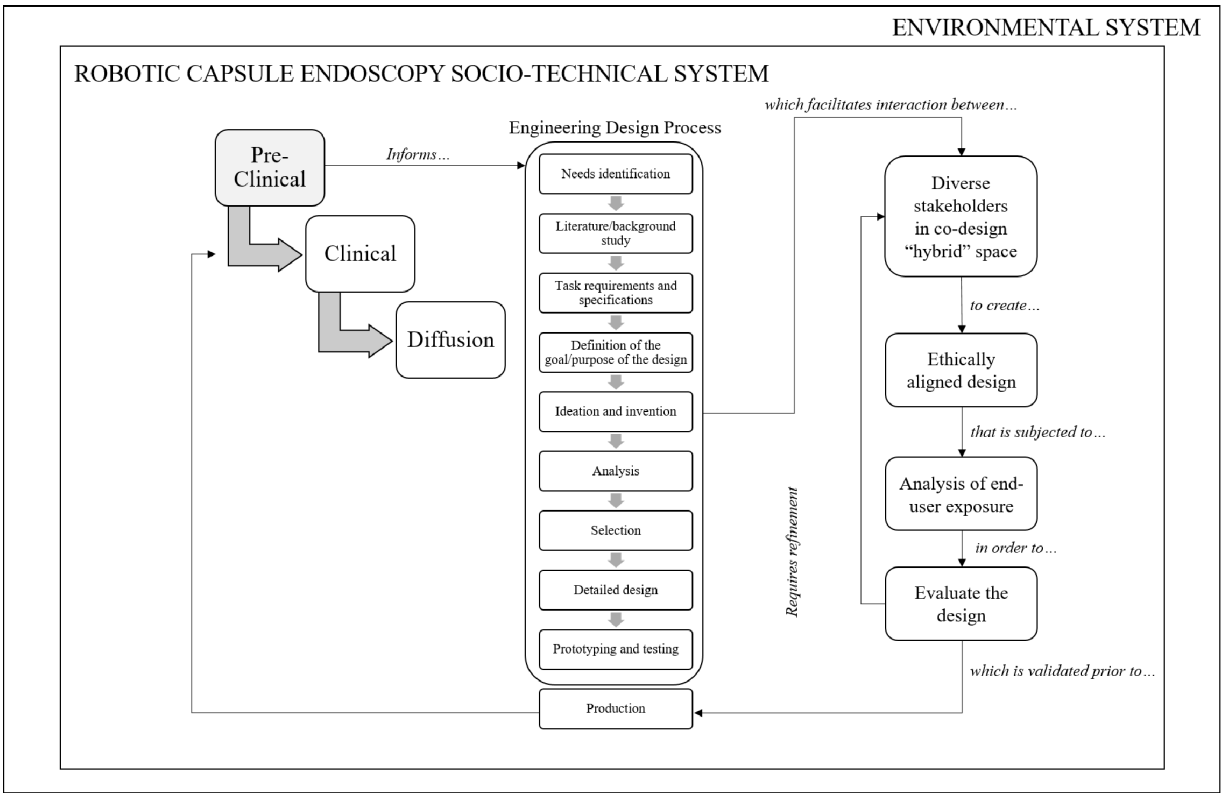

Fig. 2. ***Proposed ethically aligned co-design methodology for robotic capsule endoscopy.***

strategic level considerations informed largely by environmental influences, tactical tasks involving the actual implementation of relevant codes and standards into future capsule prototype, and finally, operational level design activities that examine end-user exposure when used. Relevant stakeholders should be approached at each level of design to avoid an isolated engineering design process.

5) End-user and additional stakeholder involvement should be solicited at the preclinical phase of development and be retained until the diffusion (postmarket) stage. The nature of this arrangement should be framed as a design partnership that functions at the various levels of design (refer to point 4 above), as required.
6) Stakeholders in the co-design of an ethically aligned robotic capsule include: engineers, external designers and developers, capsule manufacturers, medical practitioners, end-users, industry representatives, academia (including ethicists, technologists, biomedical experts, and engineers), funding bodies, government agencies, local authorities (e.g., the Therapeutic Goods Administration if operating in Australia), and other relevant international entities such as the U.S. Food and Drug Administration, and standards agencies.
7) Regulatory provisions, standards, and codes of relevance to robotic capsule endoscopy and biomedical developments include those identified in Sections II-F and II-G. These provisions should be factored into the exposure evaluation, to ensure the design is ethically aligned while also meeting specified functional requirements.
8) The theory of exposure should be used as a means to evaluate an end-user's vulnerability to an (autonomous) robotic capsule, with the intention of generating feedback about weaknesses in design, relative to a controlled version of the same technology. Such comparisons can subsequently be used to inform future design choices. Therefore, an analysis of controlled and autonomous robotic capsules is necessary.

A diagrammatic representation of the proposed co-design framework is offered in Fig. 2. The approach is generalizable to other engineering design contexts, as the progression from ethically aligned co-design to exposure analysis and redesigns, can be applied to other settings.

This paper does not commit to co-design tools and techniques that are available and could be implemented. Rather the authors focus on the need for the co-design process to provide an outcome that can be evaluated by exposure analysis, as set out below in the case study. Subsequently, a redesign process can ensue as required depending on the results. It should be noted that a design approach that may be applied in this instance in order to

determine end-user and other stakeholder values is value sensitive design (VSD).

VSD is three-tiered methodology centered on the conceptual, empirical and technological assessment of "human values" as they pertain to engineering design efforts (see [64] and [65]). While the selection of a specific design methodology is beyond the scope of this paper, existing methodologies can certainly be applied at a later stage, which may include other co-design approaches such as participatory design. However, if a design approach such as VSD were to be deployed, caution must be exercised to ensure that the methodology is: 1) responsive to cultural and situational contexts; 2) mindful of the problems associated with the definition of "universal" human values; 3) appreciative of the significance of participation, making a conscious effort to overcome potential limitations of VSD in terms of stakeholder engagement; 4) cognizant of the tradeoff among competing values and subsequent influences on socio-technical notions, such as "joint optimization"; and 5) vigilant in terms of managing the motivations of those responsible for facilitating the value-based (co-)design process. Such an approach ensures that the critiques often associated with VSD (refer to [66] and [67]) are accounted for, and that exposure can then be seamlessly integrated into the selected co-design framework.

C. Other Considerations: Limitations and Future Work

The following considerations should be noted with regards to the proposed methodology.

1) This paper is conceptual in nature, and presents an ethically aligned co-design methodology in terms of its major elements, assumptions, and considerations.
2) This paper does not claim that the co-design framework has or will be adopted entirely. Rather, this paper proposes future work should focus on stakeholder inclusivity in defining exposure and system risks that must directly be factored into future co-design of the robotic capsule and other engineering design efforts of a similar nature.
3) The scope of design evaluation is confined to the analysis of exposure, as a fundamental component of the design and development processes. This does not imply that exposure is the only means of design evaluation in a co-design effort, and nor does it deprecate the professional efforts of the "risk management" discipline as not relevant. However, the co-design methodology requires the deployment of an exposure analysis as a means of testing the design that has been produced.

IV. CASE STUDY—MODERN ENDOSCOPY AND ROBOTIC CAPSULE ENDOSCOPY

A. Current and Projected Capability

This paper considers two case studies based on endoscopic capsules: the first study uses the robotic capsule endoscopy described by [68] and [81]. This example has no autonomy, but is basically a remotely controlled robot in which all interpretation of (visual) information, and decisions to halt progress or dispense medication are solely at the discretion of the medical practitioner. For that initial case study, although the capsule is technically sophisticated, it makes no autonomous decisions and the ethical aspects of its design are those applied to any component or structure, and can hence be considered under the engineering design literature described by authors such as [2] and [69]–[72]. Advanced drug delivery systems are now in development [73], and image recognition system progress is reported in [74]. The second case study considers a hypothetical future capsule that will allow autonomous external computing systems to make diagnoses and decisions to halt progress and to administer medication autonomously. The hypothetical design would correspond to levels 8, 9, or 10 on the scale of autonomy proposed by [10] and would involve the actions of a fully moral agent as defined by [11]. The authors of this paper are aware that the predominant advantage of the endoscopic capsule is the avoidance of surgical intervention. Such avoidance is achieved without requiring autonomy, and without factoring economic drivers for the development of the capsule proposed for the second case study. However, future functional improvements may include the capability to automatically consult large electronic knowledge repositories with a similar level of accuracy to that of a medical practitioner. Since the second case study capsule makes autonomous decisions, ethical responsibilities rest with the capsule internal systems during operation, and with the designers that devolve ethical autonomy.

For the specific example of a medical device, there are a range of ethical issues unrelated to the device, which may be faced by the user, for example, "am I personally justified in spending my children's inheritance on this treatment." These issues are partially addressed by the co-design approach, but are aspects that remain outside the scope of this paper.

1) Case Study System Definitions: The authors present the subsystems of the two case studies as follows.

The proposed capsule technology described in Table 2 is largely derived from that quoted by Munoz *et al.* [68] and Mapara and Patravale [73]. The current assumption is that medication release will be in the form of liquid and not injection as there is no need to consider use of injection needles as deployment and retraction mechanisms. The precise format of the motion control system need not be specified for the purposes of this paper.

B. Exposure Analysis for Selected Case Studies

The exposure contributed by the various subsystems of the two case studies under the PMO is evaluated as follows.

1) *For the Controlled Capsule's Image System:* The battery, LED illumination, CMOS camera chip and

Table 2 Case Study Subsystems

Function	Controlled version	Autonomous ethical robot
IMAGE. When image of adequate quality for diagnosis is captured and available.	Battery, LED illumination, CMOS camera chip and lens, data compression algorithm, frame selection system (ANN), ultra-low-power transmitter, Surgeon's receiver, power supply, software and monitor	Battery, LED illumination, CMOS camera chip and lens, data compression algorithm, frame selection system (ANN). Images are not transmitted.
MOTION CONTROL. The motion control system delivers its service when the capsule's progression (under normal peristaltic motion) is halted, without damage to the tissue, and resumed at the times when start and stop actions are requested.	Surgeon's control (switch) Control transmitter and power supply external to patient. Receiver and power supply (battery) in capsule, command decoding system, actuator is the magnetic link between the target (the capsule in the gastrointestinal tract) and the external magnetic source controlled by the surgeon's switch	Sense data from IMAGE system, "Robot controller", via ultra-low-power transmitter to external receiver, internet connection to remote diagnosis system, external transmitter plus power supply to transmit control signal. Battery, receiver and decoder in capsule to generate digital result. "Action signal" generated by "consequence engine" (comprising safety ethic logic layer and "action evaluator") from digital result of diagnosis. Battery, solid-state switch, electric motor and geared actuator for "legs", with spring retraction.
MEDICATION RELEASE. The drug release system completes its function when the designed quantity of drug is delivered, without damage to the tissue, at the time/duration when the action is requested.	Surgeon's control (switch) Control transmitter and power supply external to patient. Receiver and power supply (battery) in capsule, command decoding system, actuator for medication release	Sense data from IMAGE system, "Robot controller", via ultra-low-power transmitter to external receiver, internet connection to remote diagnosis system, external transmitter plus power supply to transmit control signal. Battery, receiver and decoder in capsule to generate digital result. "Action signal" generated by "consequence engine" (comprising safety ethic logic layer and "action evaluator") from digital result of diagnosis: battery, electric motor and geared actuator for medication release.

lens, data compression algorithm, frame selection system [artificial neural network (ANN)], and ultralow-power transmitter all contribute to the E_1 single point of failure (SPOF) exposure of the system, as does the medical practitioner's receiver. The medical practitioner also contributes to the SPOF value but redundancies are likely to be available for the practitioner's power supply, software and monitor and hence these will only contribute to the E_2 value.

2) *For the Controlled Capsule's Motion Control System:* The capsule's actuator, command decoding system, receiver, and power supply all contribute to E_1. The medical practitioner's transmitter also contributes to the E_1 value but substitutes could be expected to be available and be considered to only contribute to the E_2 value of exposure.
3) *For the Controlled Capsule's Medication Release System:* The medical practitioner actually contributes to the E_1 exposure value. The practitioner's control and control transmitter also contribute to the E_1 value. However, since the external power supply and control switch are likely to have alternatives available, these are more likely to only contribute to the E_2 values. The receiver and battery within the capsule, the command decoding system and the actuator for medication release within the capsule all contribute to E_1 exposure levels.
4) *For the Autonomous Capsule's Image System:* The battery, LED illumination, CMOS camera chip and lens, data compression algorithm, frame selection system (ANN), and ultralow-power transmitter all contribute to the E_1 (SPOF) exposure of the system. The medical facility's receiver and Internet

connectivity contribute to the exposure metric/measurement, and the remote image analysis and action-recommendation system contribute considerably to the total exposure.

5) *For the Autonomous Capsule's Motion Control System:* All exposure associated with the imaging system contributes to the total exposure. The medical facility's Internet connectivity constitutes a large suite of exposure points, and the remote image analysis and action-recommendation system contribute considerably to the total exposure as these generated the basic signal recommending motion halt. The medical facility's transmitter (signal initiated by remote diagnostic system) contributes to the E_1 value. However, the medical facility's power supply is likely to contribute only to lower levels of exposure. The battery, receiver and decoder within the capsule generate a digital result and these result in a signal if the SEL and action evaluation concludes that the signal meets ethical standards [12]. The solid-state switch, electric motor and geared actuator for motion of the capsule system also contribute to the E_1 exposure value.
6) *For the Autonomous Ethical Robot's Medication Release System:* All of the exposure associated with the imaging system contributes to the total exposure. The medical facility's Internet connectivity contributes to a large suite of exposure points, and the remote image analysis and action-recommendation system contribute significantly to the total exposure. These generate the basic signal recommending a dispensation of medication. The medical facility's transmitter (signal initiated by remote diagnostic system) contributes to the E_1 value. However, the medical facility's power supply is likely to only contribute to lower levels of exposure. The battery, receiver and decoder within the capsule generate a digital result and this results in an "Action signal" if the "consequence engine" (comprising SEL and "action evaluator" using the terminology of [12], concludes that the signal meets ethical standards. The solid-state switch and actuator for the capsule's medication release system also contribute to the E_1 exposure value.

C. Ethical Considerations: Scope of Relevant Issues

It is important to consider the scope of action of both case study systems, referencing the ALFUS mission complexity and environmental complexity dimensions. For the future endoscopy capsule, the actions possible are "dispense medication" and "halt progress," and the device travels in a multidimensional trajectory in the human gastrointestinal tract. If progress is not halted, the imaging system fails and medication is not released, the capsule will pass through the patient. There are ethical considerations for the designer that are independent of the capsule's level of autonomy. For example, the motion control system can cause internal damage, either directly or through a failure to retract, and subtler questions of whether the imaging system can transmit images that will allow valid decisions. The ethical decisions for each case study are whether or not to halt progress or to release medication. For the controlled system, these decisions rest with the medical practitioner, but for the autonomous system, these decisions devolve to two sources: the treatment system and any ethical constraint system incorporated into the autonomous capsule controller.

For the autonomous endoscopic capsule, and using the arguments in [12], it would be valid to consider whether within its constrained operational environment, the autonomous endoscopic capsule could assess the most ethical option. Given an assessment, the capsule could consider whether a visualized site requires additional observation (halt progress) and whether a visualized site required administration of medication. This is not a trivial issue: a designer could apply an adaptation of the test proposed by Turing [75], expressed as "could the autonomous device pass the examinations (within specified scope, and with allowance for communication mode) that a human medical person would need to pass before being allowed to carry out the procedure."

For the particular case of the autonomous endoscopic capsule, the designer might ask whether an action evaluator and an SEL layer as proposed by Dennis *et al.* [12] is required. Fig. 1 in [12] does not explicitly address the validity of locating the SEL remotely from the robot and allowing the SEL to transmit action selections to the robot. Noting the likely complexity of a practical SEL, however, such an approach contributes considerable exposure to malicious agents beyond a designer's control, and, therefore, raises additional ethical issues.

Notwithstanding the question of whether it is ethical to locate a safety constraint system remotely from an autonomous system, the designer must consider whether it is feasible to construct and program an "action evaluator" and a safety logic layer. For an autonomous system of the complexity used in [12]'s case study and indeed that used in this paper's case studies, it seems feasible.

The practical guideline is thus proposed to be: the designer is responsible to first define the environmental complexity and the mission complexity levels (as per the ALFUS terminology). Subsequently, the designer uses a modified version of the test proposed by Turing [75] to demonstrate that the SEL is capable of achieving ethical decision making. Such decisions should be indistinguishable from that of a human as a full ethical agent who is constrained by broadly accepted criteria such as the golden rule [76] and the Universal Declaration of Human Rights [77], both of which are somewhat aligned with duty and rights-based ethics (i.e., Kantian ethics).

D. Application of the Proposed Methodology

Allowing for the exposure of the controlled endoscopic capsule used on a patient, two possible outcomes can be considered: 1) whether it will pass through the system without damage and 2) whether it will return adequate images. The processes responsible for each outcome can be presented and the exposure metric and exposure points of each can be evaluated.

For the proposed endoscopic capsule, the first outcome is identical, but the process must now model the mechanism that enables the capsule's progress to be halted and restarted without damage to the intestinal tract. The second outcome is unchanged across the two versions of the endoscopy capsule and the process itself is unchanged as long as the images are transmitted to the observer. With the addition of the capability for releasing medication, another model must be constructed and its exposure calculated. If either medication release or progress-halt can be initiated without operator intervention (i.e., resulting from capsule decisions based on internal processing of images and positioning sensors), the potential for internal decisions that have ethical implications arises and will be reflected in changes to exposure. If the progress-halt or medication-release decision making process have external connections, these will also incur considerable exposure as a result of the contributory system.

The co-design process will generate a high-level design of an autonomous capsule that meets the functional requirements of all parties. The analysis of the ethical decision making capacity has shown that for the proposed autonomous capsule case study, the ethical decisions are sufficiently constrained that it is likely to be possible to create an ethical protection layer within the robot, and, therefore, the analysis of ethical capacity has not required constraint of the level of autonomy that was proposed within the original design.

E. Design Features Arising From Methodology Application

1) Design Examination: The authors propose that the key issue for design of autonomous systems engaged in functions with ethical implications is relatively straightforward. It is important to progressively constrain the autonomous system's total scope of decision making (that has ethical obligations or significance), until the scope of action at which the proposed autonomous system can be demonstrated to be capable of decisions similar to those of an ethical human within the known environment. Furthermore, ensure that the design of the system thus scoped has a level of exposure, as defined by Robertson [58] no larger than a nonautonomous system.

2) Exposure Analysis: The analysis of exposure has drawn attention to a number of issues that apply to the controlled capsule, as well as those applicable to the autonomous capsule. In particular, the medical practitioner may either lack training or be impaired in some way, while a perfectly functional remote diagnostic system may perform better.

The power supply for the endoscopic capsule is noted as a contributor to E_1 exposure value: the use of inductive power transfer noted by Mapara and Patravale [73] can reduce exposure via redundant capacity available in the medical facility. The authors would also note the option of using either pulse code modulation or frequency shift keying of the inductive power source as a means of signaling the endoscopic capsule to reduce exposure.

Inclusion of the capability to communicate with automated diagnostic services is attractive, and seems likely that diagnostic ability equal to most medical practitioners may be available in such systems. Nevertheless, works such as [58] draw attention to the considerable level of exposure that is contributed to a system by such a connection. This represents an ethical problem given the potential for malicious use is largely outside the capacity of the designer to control.

V. CONCLUSION

The design approach proposed in this paper has been applied to case studies and shown to generate both valuable approaches to initial design and scope limitation that can be demonstrated to meet ethical requirements, and design verification processes that can ensure that the design performance is achieved even under adverse conditions.

The co-design methodology can be applied to engineering design innovations. Corporations, public agencies, universities, and research institutions have relied upon traditional methodologies driven by time to market style metrics and goals of competitive advantage. Although useful in many situations, these methodologies do not adequately account for the importance of employing the user early on in the design process. In addition, these traditional methods fail to incorporate the social and technical design elements, their interactions and the environmental factors that affect stakeholders and their engagement. Most importantly, these methodologies fail to include the impact of innovation use upon the end-user throughout the design process.

End-user and stakeholder engagement are often disregarded in engineering design endeavors, particularly in early design phases (i.e., preclinical phase in the case of biomedical engineering). Most progressive IT companies tout their ability to engage in experimental development lifecycles with agile programming techniques, as opposed to purported "outdated" waterfall systems development lifecycles. Others believe deeply in the philosophy of build it and they will come. But, is this ethical? The underlying premise of co-design, specific to biomedical device innovation, is that a problem should first exist, and a human requirement should demand that a needs identification process be initiated. This is no less relevant when

considering existing biomedical device systems from a nonautonomous to a fully autonomous setting.

A scope for generalizing any ethical approach to design is of significant concern. Over very long timeframes, the human race has achieved moderate agreement on simple ethical principles, some of which are derived from Kant's categorical imperative and the Confucian golden rule. As automation becomes more ubiquitous, society will have to truly grapple with the issue that an increasing level of ethical capability of a robot will result in an increase in its perception of self-worth. As a consequence, the more parlous could be its capability to evaluate instructions by a human and choose to ignore these instructions, if it is likely to result in a robot experiencing harm. Nevertheless, the co-design principle is demonstrated to be generally applicable, with a sufficient framework to classify a level of autonomy and situational complexity. The authors have proposed a generalizable test for constraining scope of autonomy such that embedded ethical systems can be shown to meet a standard. Finally, the authors have proposed a generalizable design approach that will ensure a given design meets acceptable criteria for end-user exposure. ■

REFERENCES

[1] K. J. Lee, Y. H. Ki, J. S. Cheon, G. Hwang, and H. S. Ahn, "Approach to functional safety-compliant ECU design for electro-mechanical brake systems," *Int. J. Automot. Technol.*, vol. 15, no. 2, pp. 325–332, 2014.

[2] J. K. H. Chan, "Design ethics: Reflecting on the ethical dimensions of technology, sustainability, and responsibility in the Anthropocene," *Des. Stud.*, vol. 54, pp. 184–200, Jan. 2018.

[3] SAE On-Road Automated Vehicle Standards Committee, "Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems," SAE Tech. Paper J3016_201401, 2014.

[4] W. Nai, Y. Chen, Y. Yu, F. Zhang, D. Dong, and W. Zheng, "Fuzzy risk mode and effect analyasis based on raw driving data for pay-how-you-drive vehicle insurance," in *Proc. IEEE Int. Conf. Big Data Anal. (ICBDA)*, Hangzhou, China, Mar. 2016, pp. 1–5.

[5] J. Stilgoe, R. Owen, and P. Macnaghten, "Developing a framework for responsible innovation," *Res. Policy*, vol. 42, no. 9, pp. 1568–1580, 2013.

[6] S. Spiekermann, "IEEE P7000—The first global standard process for addressing ethical concerns in system design," *Proceedings*, vol. 1, no. 3, p. 159, 2017.

[7] E. H. Kessler and A. K. Chakrabarti, "Innovation speed: A conceptual model of context, antecedents, and outcomes," *Acad. Manage. Rev.*, vol. 21, no. 4, pp. 1143–1191, 1996.

[8] G. Stalk, "Time-the next source of competitive advantage," *Harvard Bus. Rev.*, vol. 66, pp. 41–51, 1988.

[9] N. McBride and R. R. Hoffman, "Bridging the ethical gap: From human principles to robot instructions," *IEEE Intell. Syst.*, vol. 31, no. 5, pp. 76–82, Sep./Oct. 2016.

[10] T. B. Sheridan and W. L. Verplank, *Human and Computer Control for Undersea Teleoperators*. Cambridge, MA, USA: MIT Press, 1978.

[11] J. H. Moor, "The nature, importance, and difficulty of machine ethics," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 18–21, Jul./Aug. 2006.

[12] L. A. Dennis, M. Fisher, and A. F. T. Winfield, "Towards verifiably ethical robot behaviour," in *Proc. AAAI Workshop Artif. Intell. Ethics, Papers*, 2015, pp. 45–52.

[13] E. B.-N. Sanders and P. J. Stappers, "Co-creation and the new landscapes of design," *CoDesign*, vol. 4, no. 1, pp. 5–18, 2008.

[14] NSW Council of Social Service. (Mar. 2018). *Principles of Co-Design*. [Online]. Available: https://www.ncoss.org.au/capacity-building/sector-support/templates-and-resources/principles-of-co-design

[15] I. Burkett. (2014). An introduction to co-design. Saatavissa. Accessed: Oct. 21, 2015. [Online]. Available: http://design4socialinnovation.com.au/wp-content/uploads/2014/09/An-Introduction-to-Co-Design-by-Ingrid-Burkett.pdf

[16] E. Blomkamp, "Co-design for government: Magic bullet or magical thinking?" in *Proc. 3rd Int. Conf. Public Policy (ICPP3)*, Singapore, 2017, pp. 1–25.

[17] A. Druin, J. Stewart, D. Proft, B. Bederson, and J. Hollan, "KidPad: A design collaboration between children, technologists, and educators," in *Proc. ACM SIGCHI Conf. Hum. Factors Comput. Syst.*, 1997, pp. 463–470.

[18] A. Druin, B. Bederson, A. Boltman, A. Miura, D. Knotts-Callahan, and M. Platt, "Children as our technology design partners," in *The Design of Children's Technology*, A. Druin, Ed. San Francisco, CA, USA: Morgan Kaufmann, 1998.

[19] A. Druin, "The role of children in the design of new technology," *Behav. Inf. Technol.*, vol. 21, pp. 1–25, 2002.

[20] A. Farber, A. Druin, G. Chipman, D. Julian, and S. Somashekher, "How young can our technology design partners be?" in *Proc. PDC*, 2002, pp. 272–277.

[21] M. L. Guha, A. Druin, G. Chipman, J. A. Fails, S. Simms, and A. Farber, "Mixing ideas: A new technique for working with young children as design partners," in *Proc. Conf. Interact. Design Children, Building Community*, 2004, pp. 35–42.

[22] M. L. Guha, A. Druin, G. Chipman, J. A. Fails, S. Simms, and A. Farber, "Working with young children as technology design partners," *Commun. ACM*, vol. 48, no. 1, pp. 39–42, 2005.

[23] M. L. Guha, A. Druin, and J. A. Fails, "Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design," *Int. J. Child-Comput. Interact.*, vol. 1, no. 1, pp. 14–23, 2013.

[24] A. Albu, K. Lindberg, and A. Szymaszek, "Participatory design with preschool children: A smartphone application concept," in *Proc. SIDER*, Stockholm, Sweden, 2014.

[25] T. Bratteteig and I. Wagner, *Disentangling Participation: Power and Decision-Making in Participatory Design*. Springer, 2014.

[26] A. Freudenthal, T. Stüdeli, P. Lamata, and E. Samset, "Collaborative co-design of emerging multi-technologies for surgery," *J. Biomed. Informat.*, vol. 44, no. 2, pp. 198–215, 2011.

[27] J. Chisholm. (Mar. 2018). *What is Co-Design?* [Online]. Available: http://designforeurope.eu/what-co-design

[28] M. J. Muller and A. Druin, "Participatory design: The third space in HCI," in *Human-Computer Interaction Handbook*, J. Jacko, Ed., 3rd ed. Boca Raton, FL, USA: CRC Press, 2012, pp. 1125–1154.

[29] J. Gärtner and I. Wagner, "Mapping actors and agendas: Political frameworks of systems design and participation," *Hum.-Comput. Interact.*, vol. 11, no. 3, pp. 187–214, 1996.

[30] F. Kensing and J. Blomberg, "Participatory design: Issues and concerns," *Comput. Supported Cooperat. Work*, vol. 7, nos. 3–4, pp. 167–185, 1998.

[31] E. Björgvinsson, P. Ehn, and P.-A. Hillgren, "Participatory design and 'democratizing innovation,'" in *Proc. 11th Biennial Participatory Design Conf.*, 2010, pp. 41–50.

[32] S. Bødker, P. Ehn, D. Sjögren, and Y. Sundblad, "Co-operative design—Perspectives on 20 years with 'the Scandinavian IT design model,'" in *Proc. NordiCHI*, 2000, pp. 22–24.

[33] F. S. Visser, P. J. Stappers, R. van der Lugt, and E. B.-N. Sanders, "Contextmapping: Experiences from practice," *CoDesign*, vol. 1, no. 2, pp. 119–149, 2005.

[34] L. von Bertalanffy, "The theory of open systems in physics and biology," *Science*, vol. 111, no. 2872, pp. 23–29, 1950.

[35] F. Emery and E. L. Trist, "The causal texture of organizational environments," *Hum. Relations*, vol. 18, no. 1, pp. 21–32, 1965.

[36] L. Skyttner, *General Systems Theory: Problems, Perspectives, Practice*. Singapore: World Scientific, 2005.

[37] A. Cherns, "The principles of sociotechnical design," *Hum. Relations*, vol. 29, no. 8, pp. 783–792, 1976.

[38] A. Cherns, "Principles of sociotechnical design revisted," *Hum. Relations*, vol. 40, no. 3, pp. 153–162, 1987.

[39] F. Emery, "Designing socio-technical systems for 'greenfield' sites," *J. Occupational Behav.*, vol. 1, no. 1, pp. 19–27, 1980.

[40] E. Mumford, "A socio-technical approach to systems design," *Requirements Eng.*, vol. 5, no. 2, pp. 125–133, 2000.

[41] E. Mumford, "The story of socio-technical design: Reflections on its successes, failures and potential," *Inf. Syst. J.*, vol. 16, no. 4, pp. 317–342, 2006.

[42] E. L. Trist, "The evolution of socio-technical systems: A conceptual framework and an action research program," *Ontario Qual. Work. Life Center*, vol. 2, pp. 1–67, 1981.

[43] W. Pasmore, C. Francis, J. Haldeman, and A. Shani, "Sociotechnical systems: A North American reflection on empirical studies of the seventies," *Hum. Relations*, vol. 35, no. 12, pp. 1179–1204, 1982.

[44] W. Royce, "Managing the development of large software systems," in *Proc. IEEE WESCON*, 1970, pp. 1–9.

[45] B. W. Boehm, "A spiral model of software development and enhancement," *Computer*, vol. 21, no. 5, pp. 61–72, May 1988.

[46] B. S. Blanchard, W. J. Fabrycky, and W. J. Fabrycky, *Systems Engineering and Analysis*, 5th ed. London, U.K.: Pearson, 2014.

[47] I. Kant, *Grounding for the Metaphysics of Morals: With on a Supposed Right to Lie Because of Philanthropic Concerns*. London, U.K.: Hackett, 1993.

[48] R. Freedman, *Confucius: The Golden Rule*. New York, NY, USA: Scholastic Inc., 2002.

[49] J. P. Sullins, "When is a robot a moral agent?" *Int. Rev. Inf. Ethics*, vol. 6, 2006.

[50] H.-M. Huang, *Autonomy Levels for Unmanned Systems (ALFUS) Framework: Terminology Version 2.0*, vol. 1011. NIST, 2004.

[51] A. F. T. Winfield, C. Blum, and W. Liu, "Towards an ethical robot: Internal models, consequences and ethical action selection," in *Advances in Autonomous Robotics Systems*. 2014, pp. 85–96.

[52] K. W. Miller, M. J. Wolf, and F. Grodzinsky, "This 'ethical trap' is for roboticists, not robots: On the issue of artificial agent ethical decision-making," *Sci. Eng. Ethics*, vol. 23, no. 2, pp. 389–401, 2017.
[53] M. Vagia, A. A. Transeth, and S. A. Fjerdingen, "A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?" *Appl. Ergonom.*, vol. 53, pp. 190–202, Mar. 2016.
[54] Y. Y. Haimes, *Risk Modeling, Assessment, and Management*. Hoboken, NJ, USA: Wiley, 2015.
[55] D. J. Hand, *Measurement Theory and Practice: The World Through Quantification*. London, U.K.: Edward Arnold, 2004.
[56] K. A. Ouedraogo, J. Beugin, E.-M. El-Koursi, J. Clarhaut, D. Renaux, and F. Lisiecki, "Toward an application guide for safety integrity level allocation in railway systems," *Risk Anal., Int. J.*, vol. 38, no. 8, pp. 1634–1655, 2018.
[57] A. K. Saberi, E. Barbier, F. Benders, and M. van den Brand, "On functional safety methods: A system of systems approach," in *Proc. Annu. IEEE Int. Syst. Conf. (SysCon)*, Apr. 2018, pp. 1–6.
[58] L. J. Robertson, "Identifying and reducing technological contributions to end-user vulnerability," Ph.D. dissertation, School Comput. Inf. Technol., Fac. Eng. Inf. Sci., Univ. Wollongong, Wollongong, NSW, Australia, 2017.
[59] L. Robertson, A. M. Aneiros, and K. Michael, "A theory of exposure: Measuring technology system end user vulnerabilities," in *Proc. IEEE Int. Symp. Technol. Soc. (ISTAS)*, Sydney, NSW, Australia, Aug. 2017, pp. 1–10.
[60] L. A. Cox, Jr., "What's wrong with risk matrices?" *Risk Anal., Int. J.*, vol. 28, no. 2, pp. 497–512, 2008.
[61] L. C. Abbott and A. D. Maynard, "Exposure assessment approaches for engineered nanomaterials," *Risk Anal., Int. J.*, vol. 30, no. 11, pp. 1634–1644, 2010.
[62] C. Perrow, *Normal Accidents: Living with High-Risk Technologies*. New York, NY, USA: Basic Books, 1984.
[63] L. J. Robertson, K. Michael, and A. Munoz, "Assessing technology system contributions to urban dweller vulnerabilities," *Technol. Soc.*, vol. 50, pp. 83–92, Aug. 2017.
[64] B. Friedman, P. Kahn, and A. Borning, "Value sensitive design: Theory and methods," Univ. Washington, Seattle, WA, USA, Tech. Rep., 2002, pp. 2–12.
[65] B. Friedman, "Value-sensitive design," *Interactions*, vol. 3, no. 6, pp. 16–23, 1996.
[66] J. Davis and L. P. Nathan, "Value sensitive design: Applications, adaptations, and critiques," in *Handbook of Ethics, Values, and Technological Design*. Springer, 2015, pp. 11–40.
[67] N. Manders-Huits, "What values in design? The challenge of incorporating moral values into design," *Sci. Eng. Ethics*, vol. 17, no. 2, pp. 271–287, 2011.
[68] F. Munoz, G. Alici, and W. Li, "A review of drug delivery systems for capsule endoscopy," *Adv. Drug Del. Rev.*, vol. 71, pp. 77–85, May 2014.
[69] M. Hersh, "Ethical engineering: Definitions, theories and techniques," in *Ethical Engineering for International Development and Environmental Sustainability*. London, U.K., 2015.
[70] S. P. Nichols, "Professional responsibility: The role of the engineer in society," *Sci. Eng. Ethics*, vol. 3, no. 3, pp. 327–337, 1997.
[71] C. B. F. Leddermann, *Engineering Ethics*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2012.
[72] H. Elder, "Discussion: Ethics and the professional engineer," in *Transactions of the Institution of Professional Engineers New Zealand: General Section*, vol. 16. 1989.
[73] S. S. Mapara and V. B. Patravale, "Medical capsule robots: A renaissance for diagnostics, drug delivery and surgical treatment," *J. Controlled Release*, vol. 261, pp. 337–351, Sep. 2017.
[74] Y. Kominami *et al.*, "Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy," *Gastrointestinal Endoscopy*, vol. 83, no. 3, pp. 643–649, 2016.
[75] A. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, pp. 433–460, 1950.
[76] R. Haupt and A. J. Shockley, "Golden versus platinum rules [ethically speaking]," *IEEE Antennas Propag. Mag.*, vol. 59, no. 4, p. 118, Aug. 2017.
[77] United Nations General Assembly in Paris. (Mar. 26, 2018). *General Assembly Resolution 217 A: Universal Declaration of Human Rights*. [Online]. Available: http://www.un.org/en/universal-declaration-human-rights/
[78] L. McIntyre, K. Michael, and K. Albrecht, "RFID: Helpful new technology or threat to privacy and civil liberties?" *IEEE Potentials*, vol. 34, no. 5, pp. 13–18, Sep./Oct. 2015.
[79] K. Michael, "Can good standards propel unethical technologies?" *IEEE Technol. Soc. Mag.*, vol. 35, no. 3, pp. 6–9, Sep. 2016.
[80] K. Michael, "Implantable medical device tells all: Uberveillance gets to the heart of the matter," *IEEE Consum. Electron. Mag.*, vol. 6, no. 5, pp. 107–115, Oct. 2017.
[81] F. Munoz, G. Alici, H. Zhou, W. Li, and M. Sitti, "Analysis of magnetic interaction in remotely controlled magnetic devices and its application to a capsule robot for drug delivery," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 298–310, Feb. 2018.

ABOUT THE AUTHORS

Lindsay J. Robertson received the B.E. degree in mechanical engineering design and thermal systems from Canterbury University, Canterbury, New Zealand, in 1976, the M.Tech. degree (honors) from Massey University, Palmerston North, New Zealand, in 1990, and the Ph.D. degree from the University of Wollongong, Wollongong, NSW, Australia, in 2017, with a focus on the theme of technological risk, exposure, and resilience.

From 1976 to 1987, he held positions with the New Zealand Government. From 1990 to 2007, he was with Fonterra (and NZ Dairy) Research Centre, New Zealand. From 2007 to 2016, he was a Principal Engineer with Parsons Brinckerhoff, New Zealand.

Dr. Robertson has been a Fellow within the Institution of Professional Engineers in New Zealand (IPENZ) since 1999 and also within the Institution of Mechanical Engineers (U.K.) since 2013. He was the Editor-in-Chief of *IPENZ Transactions* from 2002 to 2016.

Roba Abbas received the B.S. degree (with first class honors and distinction) in information and communication technology (business information systems) and the Ph.D. degree in location-based services regulation from the University of Wollongong, Wollongong, NSW, Australia, in 2006 and 2012, respectively.

From 2005 to 2010, she was a Product Manager with Internetrix, Wollongong. Since 2013, she has been an Honorary Fellow with the School of Computing and Information Technology, University of Wollongong, Wollongong. Since 2017, she has been a Sessional Lecturer with the School of Electrical, Computer and Telecommunications Engineering. She is currently a Lecturer with the School of Management, Operations and Marketing, University of Wollongong. She has authored or co-authored over 25 refereed papers. Her current research interests include methodological approaches to design, including co-design and socio-technical systems.

Dr. Abbas has been the Associate Editor of *IEEE Technology and Society Magazine* since 2016. She was a recipient of the Australian Research Council Scholarship. In 2006, she earned a place on the dean's merit list. In 2012, her Ph.D. was awarded Examiners' Commendation for Outstanding Thesis.

Gursel Alici received the Ph.D. degree in robotics from the Department of Engineering Science, Oxford University, Oxford, U.K., in 1994.

He is currently a Senior Professor with the University of Wollongong, Wollongong, NSW, Australia, where he has been the Head of the School of Mechanical, Materials, Mechatronic and Biomedical Engineering since 2011. He has authored or co-authored over 300 refereed publications and delivered numerous invited seminars and keynote talks in his areas of research. His current research interests include soft robotics, system dynamics and control, robotic drug delivery systems, novel actuation concepts for biomechatronic applications, robotic mechanisms and manipulation systems, soft and smart actuators and sensors, and medical robotics.

Dr. Alici was a member of the Mechatronics National Panel formed by the Institute of Engineers, Australia from 2007 to 2017. He was a recipient of the Outstanding Contributions to Teaching and Learning Award in 2010, the Vice-Chancellor's Interdisciplinary Research Excellence Award in 2013, and the Vice-Chancellor's Award for Research Supervision in 2018 from the University of Wollongong. He has held a Visiting Professorship position at the Swiss Federal Institute of Technology, Lausanne (EPFL), City University of Hong Kong, and University of Science and Technology of China (USTC). He was a Technical Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS from 2008 to 2012. He is currently a Technical Editor of IEEE ACCESS. He has served on the international program committee of numerous IEEE/ASME International Conferences on Robotics and Mechatronics. He was the General Chair of the 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics held in Wollongong, Australia. He is the Leader of soft robotics for the Prosthetic Devices theme of the Australian Research Council Center of Excellence for Electromaterials Science from 2014 to 2021.

Albert Munoz received the B.S. degree in chemical engineering and the M.B.A. degree from the Florida Institute of Technology, Melbourne, FL, USA, in 1999 and 2001, respectively, and the M.Sc. degree in environmental engineering and the Ph.D. degree in supply chain management from the University of Wollongong, Wollongong, NSW, Australia, in 2004 and 2012, respectively.

From 2009 to 2011, he was a Research Fellow with the SMART Infrastructure Facility, University of Wollongong, where he is currently a Senior Lecturer with the Faculty of Business, School of Management, Operations & Marketing. He has authored or co-authored refereed papers in systemic outcomes to risk, including disruption management, resilience, robustness, and performance measurement.

Katina Michael (Senior Member, IEEE) received the B.S. degree in information technology from the School of Mathematical and Computing Science, University of Technology, Sydney, NSW, Australia, in 1996, the Doctor of Philosophy degree in information and communication technology from the Faculty of Informatics, University of Wollongong, Wollongong, NSW, Australia, in 2003, and the Master of Transnational Crime Prevention degree (distinction) from the Faculty of Law, University of Wollongong in 2009.

She has held visiting academic appointments at Nanjing University, Nanjing, China, and the University of Southampton, Southampton, U.K. From 1996 to 2001, she was a Senior Network Engineer with Nortel Networks, Wollongong, NSW, Australia. She was also a Systems Analyst with Andersen Consulting, North Sydney, NSW, Australia, and OTIS Elevator Company, Minto, NSW, Australia. She is currently a Professor with the School for the Future of Innovation in Society and School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA, where she is also the Director of the Center for Engineering, Policy and Society. She has authored or co-authored over seven edited books, 500 page reference volume: *Innovative Automatic Identification and Location-Based Services: from Bar Codes to Chip Implants* (Hershey, PA, USA: IGI, 2009), and 150 peer-reviewed papers. Her current research interests include emerging technologies and technologies used for national security and their corresponding social implications.

Ms. Michael was a recipient of the Brian M. O'Connell Distinguished Service Award in the Society for the Social Implications of Technology. She is the Guest Editor of 15 special issues including in the PROCEEDINGS OF THE IEEE, *Computer, IEEE Robotics and Automation Magazine*, *IEEE Potentials*, *Journal of Location-Based Services*, *Computer Communications*, *Electronic Commerce Research*, and *Prometheus*. She was the Editor-in-Chief of the *IEEE Technology and Society Magazine* from 2012 to 2017 and has been a Senior Editor of the *IEEE Consumer Electronics Magazine* since 2015. She is the founding Editor-in-Chief of the new *IEEE Technology and Society Magazine* that will begin publication in 2020.