

# Teaching Ethical Design in the Era of Autonomous and Intelligent Systems

Thomas Lee  
Quanser Consulting  
Markham, Canada  
tom.lee@quanser.com

Jennifer Boger  
Systems Design  
Engineering  
University of Waterloo  
jboger@uwaterloo.ca

Maurice Mulvenna  
School of Computing &  
Mathematics  
Ulster University  
md.mulvenna@ulster.ac.uk

Raymond Bond  
School of Computing &  
Mathematics  
Ulster University  
rb.bond@ulster.ac.uk

**Abstract**— How can the Turing-complete machines of tomorrow be designed ethically by the students of today? As the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems notes, there is growing concern that AI and other transformational technologies “remain human-centric, serving humanity’s values and ethical principles.” This concern is amplified in the context of engineering education as our curriculum still struggles to enter the 21st Century let alone address the accelerating complexities of modern era systems. This paper offers a contemporary perspective on the teaching of ethics as an integral part of design itself. Part of the framework of discourse is the notion of “ethical by design” as proposed by three of the authors of this paper, Mulvenna, Boger, and Bond. They proposed a pragmatic manifesto of principles to support developers, providers, and users in the collaborative process of inherently and explicitly including ethics into product and service design. One can view this manifesto as a practitioner’s perspective and we argue that it offers a useful starting point to revise engineering education’s historical approach to ethics as a legal and professional framework, to one that is accessible and integral in the engineering design process, adaptable to the evolving complexities of engineering.

**Keywords**—Ethical design; AI; Engineering ethics

## I. INTRODUCTION

In 2016 the IEEE, a major engineering professional society which is also known for establishing some of the most important standards of the modern digital ecosystem, launched the Global Initiative on Ethics of Autonomous and Intelligent Systems professes to “move beyond the paranoia and uncritical admiration regarding autonomous and intelligent technologies” [1]. This one simple phrase clearly illustrates the emerging paradox of this technology context. Many are thrilled at the prospect of a safer, cleaner, and even more prosperous global future via technologies that will optimally coordinate societal function. But at the same time there is a parallel conversation the potentially detrimental effects of the technology. Elon Musk has posited that AI is, in his opinion, society’s “greatest existential threat” [2]. The topic has entered mainstream conversation as a broad cross-section of the public begin to realize that we have already entered a world where arguably intelligent technology has infiltrated our lives and likely will rapidly increase its reach. Smartphone maps, social

media, smart safety features in cars, and countless other personal technologies have fundamentally changed users’ behavior and perspectives on their relationship to their environment.

## II. THE ENGINEERING PERSPECTIVE

This paper begins to address key derivative questions from the above mainstream context. As the designers, builders, and maintainers of these new technologies, is there a parallel and arguably more important discussion that we must engage in? Of course the answer is “yes” and the substantive initiative like the IEEE Global Initiative, as well as many academic, government, and public discussions and presentations on this topic are clear indications that engineers have a special role in ensuring an effective and responsible future for the emerging technology.

In many ways, these initiatives of professional conscience among global engineers is not surprising as the profession has historically had a relatively strong and purposeful engagement with ethics in a formal sense. In Canada, the home country of two of the authors, the field of engineering is a self-regulating profession that is sanctioned to provide a professional and legal framework that protects the greater public interests when professional engineers engage in their work. Part of this framework is a deliberate and formal treatment of ethics at both the national and provincial level (e.g. the profession of engineering in Canada [3][4]), the essence of which, are integrated into the professional licensing process and undergraduate programs. Indeed, engineering professional bodies and regulators consider ethics to be a key component of professional competence.

In terms of substance, the existing formal framework of ethics appear to focus on matters of existing legal frameworks (in places such as Canada where there are strong legal structures defining and governing the profession), and matters of personal character, as, for example, implied by the Canadian Code of Ethics. Neither perspective, have sufficient granularity to help engineers, and the emerging generation of engineers, to guide them in actual design practice above and beyond the attitudinal and character elements, as important as they are. Moreover, ethics is usually taught as a stand-alone

topic in a single course or short series of lectures, rather than as an integrated and integral component throughout the curriculum. Specifically, while the nature of our technologies are dramatically increasing in complexity that have potential implications, how we train engineers with respect to ethics remains relatively static.

### III. COMPLEXITY OF MODERN ERA SYSTEMS

Succinctly, it is no longer possible to predict or even define potential “failure” of systems. Current modeling techniques that we teach undergraduates imply that a relatively compact set of equations or other deterministic model forms can guide design. Realistically, modern-era systems embody complexity that are well beyond the level of fidelity implied by conventional modeling techniques.

- Modern-era components and devices are dynamically and computationally more complex than conventional objects of design.
- Components and devices are connected and can communicate and make decisions autonomously. New tools of development and testing are often “black boxes” and system-level uncertainty can be a major challenge.
- Intelligent components and devices, by definition, adapt their behavior creating a time-variant complexity that makes all designs non-deterministic.

The “nightmare scenario” in the modern era is that systems that we design, even with most robust ethical intent and application of our best instincts, can be inappropriate for their use context or go “rogue” in operation. The increasingly significant responsibility and trust we are placing in technology means failures result in increasingly and bring about disastrous unintended consequences.

### IV. IEEE: “ETHICALLY ALIGNED DESIGN”

Can the design process that we introduce to our students have inherent methods that can reduce the probability of our systems going rogue and producing resulting in unintended, harmful consequences? Currently, the most ambitious response to this question is that of the IEEE Global Initiative.

At a practical level, the Global Initiative seeks to first establish a strong formalized global dialog on key application areas where the issues of ethics are particularly more precarious and ultimately it seeks to establish a comprehensive series of key standards under the umbrella of IEEE P7000 [5].

- IEEE P7000™ - Model process for addressing ethical concerns during system design
- IEEE P7001™ - Transparency of autonomous systems

- IEEE P7002™ - Data privacy process
- IEEE P7003™ - Algorithmic bias considerations
- IEEE P7004™ - Standard on child and student data governance
- IEEE P7005™ - Standard for transparent employer data governance
- IEEE P7006™ - Standard for personal data artificial intelligence (AI) agent
- IEEE P7007™ - Ontological standard for ethically driven robotics and automation systems
- IEEE P7008™ - Standard for ethically driven nudging for robotic, intelligent, and automation systems
- IEEE P7009™ - Standard for fail-safe design of autonomous and semi-autonomous systems
- IEEE P7010™ - Wellbeing metrics standard for ethical artificial intelligence and autonomous systems

The current state of the initiative stands at the release in 2017 of the comprehensive document Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. This is the second version integrating the public feedback generated by the first version released in 2015. As yet, the actual details and recommendations in the form of the standards details are still in respective committee discussions.

The IEEE Global Initiative is comprehensive and in some sense, may be more challenging for many institutions and individual instructors to adapt to existing courses. Indeed the typical goal of any standards initiative is to provide comprehensive guidance and frameworks. Notwithstanding, the Ethically Aligned Design initiative does acknowledge a need to derive an education vision and framework. Clearly, once the work is done, it will be a welcome framework that will likely influence the evolution of properly integrating ethics into design education.

The issue is, of course, what do we do at this moment in time? In many parts of the world, robotics, and other forms of intelligent mechatronics have become core undergraduate programs and programs dedicated to artificial intelligence are emerging. Academics are pursuing ambitious research goals within intelligent system contexts without well-acknowledged frameworks. Increasingly, well-known technology companies openly admit their desire to accelerate the commercialization of intelligent technology. This high-speed train has left the station and is speeding off to potentially unknown destinations.

### V. “ETHICAL BY DESIGN”

The notion of “Ethical by Design” was introduced by Mulvenna, Boger, and Bond as a pragmatic step in the deliberate evolution of a practical ethical framework [6]. These authors argue that dimensions of ethics can be

integrated into design thinking itself and to this end, they proposed a manifesto of “ethical by design” principles that is intended to support developers, providers, and users across disciplines and sectors in the collaborative process of inherently and explicitly including ethics into product and service design.

- Design to support the people who will be using the product or service by engendering empathy for users.
- Provide enough information for people to make informed decisions at every stage about whether, when, and how to use the product or service.
- Respect people’s right to choose how they engage with the product or service; offer alternatives or customization.
- Balance appropriate privacy and security with equitable access by as many systems and people as possible, globally.
- Seek to integrate with and support the progression of policy.
- Actively look for and challenge biases and values that may be reflected in a product or service design.
- Complement differing needs, abilities, viewpoints and morals.
- Support shared decision making and feedback.
- Aim for economically, environmentally, and socially sustainable designs.
- Integrate planning for how to handle failure, including transparency and reporting.
- Be realistic about what is possible and needed.
- Support the product or service throughout its lifespan.

The “ethical by design” manifesto appears to be a unique prescriptive tool that provides another angle to ethical thinking for engineering students. The basic propositions are accessible in meaning and spirit to students and importantly, it provides a meaningful framework for vigorous dialog within the course.

As an interesting exercise, one may wish to challenge students to reflect on the design of Facebook or the smart phone and whether they embody all or most of the elements of the manifesto. These technologies have had celebrated beginnings but are now embroiled in heated public debate as to whether they are positive or negative influences on our lives. The simple act of asking the right questions is often the heart of ethics education and the manifesto offers a balance of questions that are ethically meaningful but also relate well to the technology and application context of engineering applications.

The point is, of course, these are complex issues with no clear answers and a critical part of the intellectual and professional development of engineers is his/her ability to reconcile these questions and still achieve the practical demands of their careers and their respective enterprises. In this sense, Mulvenna et al. offers these points as “sign posts”,

or directional vectors of perspective development as opposed to structured guidelines. Notwithstanding, the authors of this WEEF/GEDC paper, contend that there is sufficient structure and the key consequent questions can readily be introduced in various courses relating to design, coding, and modern-era applications of various sorts.

## VI. ETHICAL THINKING IN MODERN ENGINEERING EDUCATION

In September 2018, Tsinghua University, Beijing China, hosted the first International Forum on Engineering Education. This conference drew senior administration and professors from prominent Chinese and non-Chinese institutions that offer comprehensive engineering programs. Within the conference program was the panel session “Engineering Ethics in the Modern Era”. Two principal conclusions emerged from this session.

First was the continuing development and refinement for the formalism of engineering ethics for modern applications. For example Lee [7] proposed the integration of the core philosophical foundations of ethics – virtue, duty, community, and consequences – within engineering design thinking as a complement to the more practical aims of the manifesto of Mulvenna et al. These four pillars, according to Lee, are particular variants of philosophical categorizations of ethics that suited the context of engineering education. Other pillars often cited include autonomy, beneficence/non-maleficence, and justice. Notwithstanding, such an approach of triggering challenging discussion can promote ethical thinking during the design process, Lee contended.

Another point of vigorous discussion at the conference, was the need for a comprehensive collection relevant, contemporary case studies. The technological landscape has changed dramatically within the last two decades and engineers today must contend with systems of unprecedented complexity. Panelists from both Chinese and non-Chinese universities noted the importance of good case studies to help bring key questions and stakes into the classroom.

Currently such a collection of case studies that deal explicitly with contemporary issues and applications are not common. A notable work, however, is Tech ethics curricula: A collection of syllabi [8]. This on-line resource is a list of over 200 existing courses from around the world. All of the courses approach the topic within the framework of modern-era technology themes. Over 130 of the cited courses are at the undergraduate level. This comprehensive list is a clear indication of the growing, interdisciplinary interest in sound academic platforms to harmonize ethics thinking with engineering design thinking.

At the core, as technology continues to rapidly increase in complexity, prevalence, and autonomy, the methods to train its creators and regulate its use are becoming antiquated. We

need to train the people who build it to engage in more inherent and thoughtful ethical design. It is clear that the required academic dialog as begun in earnest.

## VII. CONCLUSIONS

As engineering embraces nascent techniques from AI and other "bleeding-edge" technologies, how we teach students to tackle the unprecedented ethical challenges become critical. This paper offers a snapshot of notable activities within various engineering and academic communities to establish a structure and pragmatic frameworks for reconciling the many vexing dimensions of contemporary ethics with engineering design. The various initiatives range from very ambitious ones such as the "Ethically aligned design" from the IEEE to more modest but arguably more practical approaches, as exemplified by the "ethical by design" work. The clearest conclusion is likely how important this issue seems to be to the many stakeholders, among which, the global engineering education community has arguably a greater obligation to establish societal leadership. Notwithstanding, ethics by tradition and in its essence, is interdisciplinary and a diverse ecosystem of vigorous dialog has already taken root.

## REFERENCES

- [1] IEEE Standards Association, "Ethically Aligned Design, Version 2", 2018.
- [2] S. Gibbs, "Elon Musk: artificial intelligence is our biggest existential threat", *The Guardian*, October 27, 2014.
- [3] Engineers Canada, "National Guidelines on the Code of Ethics", 2016.
- [4] P. Shaw, J. McPhee, G. Andrews, *Canadian Professional Engineering and Geoscience*, Nelson 2018.
- [5] IEEE Standards Association, "P7000 - Model Process for Addressing Ethical Concerns During System Design", 2016.
- [6] M. Mulvenna, J. Boger, R. Bond, "Ethical by Design – A Manifesto", *Proc. ACM-ECCE 2017*, Umeå, Sweden, 2017.
- [7] S. Lee, "Engineering Ethics in the Modern Era", oral presentation at the *International Forum on Engineering Education*, September 24-25, 2018, Tsinghua University, Beijing.
- [8] C. Fiesler, "Tech Ethics Curricula: A Collection of Syllabi", published on medium.com, 2018.