# SOCIAL AWARENESS IN DESIGN & ENGINEER EDUCATION AND PRACTICE. THE VALUE OF ETHICS IN POSTGRADUATE EDUCATION.

**Dr Laura Ferraello,**
[1]Royal College of Art

**ABSTRACT**
This paper describes how the practice of ethics and morality in design and engineering education can foster an industry which is ethically capable of tackling social issues. Through ethically aware design methodologies that postgraduate cross-discipline students developed through workshops, this paper outlines and discusses possible directions and methods that design and engineering education could evolve with further study. What can ethics awareness in postgraduate education trigger and how can it influence future design and engineering practice? For the major role morality and ethics have in ruling social coexistence designers and engineers are called to take responsibility of any action that shapes behaviour between people, products and systems. To respond to the need of a collective ethical etiquette that products and systems can encode, design research at the Royal College of Art has been supporting a practice that acknowledges impacts and people's responsibilities. A series of workshops will describe the role of collaboration and engagement in drawing ethically aware design and engineering processes; it will be evidenced how these are strategic in modulating relationships and behaviours and in guiding the mapping of the interactions of people with people and machines to acknowledge roles, responsibilities and the ownership of decision-making.
The intention of this paper is not to draw conclusions on what ethics is and how it should be practiced, but to support the development of guidelines that can engage design and engineering education and industry with debates on responsibilities and decision-making across the lifecycle of a product, system, service or infrastructure. This paper intends to be an engaged, collegial and collaborative contribution to the topic of ethics in design/engineering education and practice.

*Keywords: Ethics, design methods, collaboration, responsibility, innovation, future of education, future of practice in design and engineering*

## 1   INTRODUCTION

21st century technological development of artificial intelligence is reshaping and reorganising the traditional systems upon which society is built, both at an organizational and ontological level. People's communication and relations are predominantly looked after by algorithms, which oversee human to human interactions. The scale and granularity of this asks the revision of the role and value of technology; with 21st century advancements technology becomes "*a socio-technical assemblage*" [3] i.e. an infrastructure which assembles, reviews, distributes, organises and delivers people's relations. This doesn't come clear of issues, of which the most relevant is the devaluation of human agency. Algorithms are designed to facilitate and organise people's interaction, but by doing so they take over control of mankind relations, thus limiting the acknowledged and unacknowledged human decision-making. This disempowers people and makes them rely more heavily on technology, thus provoking a deskilling process that weakens the human capability of taking decision and understanding of short and long term effects [6]. As consequence, humans are not aware of the principles upon which decisions are made and are engaged in tasks with no sense of the impacts these can cause in the short- and long-term. Under this trend, society is increasingly becoming a black box [2]; the adoption of technology through an un-strategic planning evidences a gap of knowledge in foreseeing types and scales of social impacts. 21st century adoption of technology echoes the early 19th century one [7]. This prioritised machine efficiency upon other parameters, which impacted the

relationships between human and machines [7]. If scientists and philosophers like Watt, Babbage and Bentham attempted to design a system where women and men could fit machinic mechanisms, 21st century artificial intelligences moved humanity to the background. What is at stake with people losing cognitive control over machines is the very ontology of society, i.e. the nature of being human and act as such in a community. The process used by artificial intelligences to improve efficiency in human systems and infrastructures makes humans a component of the systems. This unbalances relationship between people and machines, as there is no competition with the technological pace and speed of change. This is an ethical concern: with machines running the pace of change humans have little space to develop strategies that support technological advancements for the benefit of society [6]. In addition, the high and relentless pace of change alienates bodies, as both Karl Marx and Luciano Floridi describe [8]; for both philosophers bodies are alienated from themselves when they lose control of the self projected on labour (Marx) and data (Floridi) [8]. Humanity has been designing infrastructures that alienate people and deprives them of the agency to understand and make decisions. This condition questions the nature of being human and the role humanity can hold in the future.

This is the context that motivates Royal College of Art (RCA) researchers to look for alternative strategies that can design a different future. With a critical and proactive acknowledgement of the ethical issues arising from human-machine-human interactions RCA researchers are looking for opportunities that can draw a different scenario. This brings to the surface ethical issues to outline actions capable of modifying the course of the events. Using the Alpha Go example, i.e. the algorithm that beat the human player, former national-security adviser and secretary of state to Presidents Richard Nixon and Gerald Henry Ford  Kissinger points out that the human capability is not about winning but thinking solutions, i.e. learning the strategy with which the machine succeeded and create something new [6]. Even though technology granularly permeates human daily life, it is important to reflect on the opportunities it can bring. What the strategy and the method could be is nonetheless the conundrum, which needs further study. One possible pathway of research is illustrated in this paper. This looks at, analyses and questions the interactions between human, human and machines through the lenses of ethics. Ethics, as practice, embodies parameters of reflection, discussion, criticality and debate, which help draw clarity and critical understanding. In human-machine-human interactions ethics can play the dual role of outlining issues, like social integration of technology, and help define strategies. In other words, ethics is both capable of identifying issues and drawing solutions in a complex area like social innovation. Ethics is also a scalable and malleable approach, as its practice is based on discursive open ended interactions; this makes ethics suitable for industries with different issues and challenges. Ethics can indeed draw cross-sector strategies that different industries can adopt and implement. In the context of this paper the focus will be limited to art, design and engineering. The following paragraphs will outline how interaction can be a methodology that applies ethics principles (discussion, refection, etc) and includes and reflects the different perspectives and behaviours of the parties involved in the development of products, systems and services. To design this methodology it is indeed necessary to outline rules, tasks and behaviours that create patterns in the system. This requires an approach that understands technology as a *socio-technical assemblage* [3] integrated in the system and capable of agency, likewise humans. Under these terms it is necessary to design a new ecology where humans and machines equally interact and exchange information [4]. This exchange is the domain of ethics. It then follows that this ecology exchange is designed and enabled by dialogue, reflection and criticality, which give humans agency of evaluating, reflecting, understanding. Interactions (or exchanges) between humans and machines across the whole systems are the output of this methodology.

This is an ambitious and challenging approach, in particular for the complexity of relationships between the roles involved in the process. To tackle this aspect RCA researchers have been developing a platform that simulates the complexity of the decision-making process under the perspective of different roles; this is to train future cross-discipline professionals to develop tools applicable to different industries. The platform is composed of two workshops which tested the ethics methodology by using design methods to facilitate dialogue across parties. These workshops aimed at fostering an ethically aware design-production-development-disposal process which is measured against the human capability to think, reflect and create. The ambition of this platform is to tackle the complexity of issues emerging from the development of product, service and infrastructure and, at the same, design tools that deal with this complexity. The workshops use mainly collaboration and co-design as epistemological purveyors of this approach. By identifying the different parties acting in the

ecosystem, these workshops intend to draw an ethical methodology which recognises the distributed nature of work 4.0, in which exchange and decision making are core and also glitch of the system. Drawing from the definition of decision-making as discursive and dialectic interaction, the workshops aim at designing a cross expertise ethically aware design-development-distribution-disposal process that can be adopted and implemented by different industries. Under these lenses 21st century ethics becomes a method that designs human-machine-human interactions to develop a critical approach that questions machine-human decision-making.

## 2   METHODS AND APPROACCHES FOR TEACHING 21ST CENTURY ETHICS METHODOLOGIES

Global contemporary education has been developing ethics courses to be deployed in STEM disciplines. These are experiments which offer some interesting points of reflection. The Harvard Embedded EthiCS, for instance, is a framework that sees teaching ethics in STEM education as a complementary factor. In this platform ethics is understood as a distributed pedagogy integrated into traditional computer science curriculum [1]. The platform aims at instilling doubts; through the integration of ethics trained researchers at a doctoral and postdoctoral level, undergraduate students are encouraged to question computer science from an ethical and moral perspective, which acknowledges responsibilities in the decision-making process. "*Students can learn to think not only about what technology they could create, but also whether they should create that technology*" [5]. Under these terms ethics becomes a guiding principle casting a different light on the topic of study. The platform trains students (1) to negotiate with ethical issues against efficiency to foster accessibility and inclusivity; (2) to understand how ethics permeates the discipline; (3) to question computer science from an ethical perspective which enables familiarisation with the complexity of a discursive approach [5]. Findings evidence that the co-design and collaboration approach helped students overcome the insecurities rising from the lack of knowledge and expertise in ethics and from the unsettling factor of being exposed to multiple and possible answers, which contrast with the traditional problem-solving approach used in computer science [5]. Through collaboration and engagement EthiCS distributed platform develops ethics from and for computer scientists. Active learning activities enable the collaborative practice, test students' understanding of ethics principles and help consolidate the bridge between the disciplines.

Collaboration is also the method two other universities used to teach ethics. The University of California (UC) Berkley and Delft University of Technology (TU Delft) developed a graduate engineering ethics pilot course for PhD students, which aimed to teach ethics across computer science and philosophy to foster cross-discipline research. The project started from questions interrogating the methods, the validity of the combined approach and how to foster collaboration [9] The pilot took a similar approach to Harvard EthiCS, as it integrated ethics courses in the existent curriculum to "*systematically addressing moral problems in engineering*" [9]. The pilot was divided in three different phases, two taking place in the respective universities, aimed at developing the methods, and a third one seeing the exchange of staff across the two partner universities. This project included in the teaching activities the awareness of students' pre-existing bias and assumptions, which impact the emotional reaction to ethical issues. As the pilot was aimed at fostering new opportunities of research across fields and countries, culture, with its biases and assumptions, became the relevant aspect that designed the success and/or failure of the project. If the exchange of staff between disciplines helped grow ownership and familiarity with the topic (the teaching was promoted by engineers who are best aware of the problems), this worked well in TU Delft only, as its academic culture normally promotes cross discipline teaching. For the same reason it didn't work in Berkley UC where cross-discipline teaching is considered second rate. The culture and readiness of learning was then pivotal for the success of the pilot; two cultures collaborating on the same project came with two different results not for the methods deployed but for the cultural environment. As pointed out by the Delft partners, methods can thrive with the right cultural environment; with ethics in particular culture is key for the role it plays on behaviour and moral. It then follows that in this area any method needs to account the diversity of cultures to design social innovation.

The context in which the RCA workshops are located benefits from an embodied interdisciplinary and intercultural context. The RCA counts four Schools (Design, Architecture, Communication and Fine Art), which body of staff and students come from any part of the globe and covers different kinds of disciplines which include, and are not limited to, design, engineering, art and humanities. The college

has a solid history in working across disciplines. This is designed at MA, MRes, PhD and MPhil level; students are driven to learn from each other. In this kind of environment collaboration is a process that is fairly familiar to the majority of the students. The same doesn't apply to ethics. The college has bodies that review the ethics of student's and staff's projects. This is practiced through a colour coded form that outlines issues. Nonetheless this procedure undermines the value of ethics, as students perceived it as a barrier to overcome and get clearance of. However, the culture of the college normally gravitates on designing social and cultural innovation through the practice of art, design and engineering and ethics is something that is latently alive across the college and at any level of study. Nonetheless, with the existent process, ethics suffers from low consideration for the procedural systems in which it is practiced. This results in disengaging students, even though ethics committees have been created to help students overcome ethical problems. The challenge is then designing a cultural context where ethics can be properly acknowledged as a component supporting opportunities across practices, rather than designing barriers. To tackle this a set of workshops have been deployed in the research department of the School of Design and in the cross-college Doctorial training to design a proactive ethically aware practice of design, art and engineering. These workshops are based on collaboration, participation, experience and co-design which are deployed to engage cross-disciplines students with the development of bottom-up ethical principles on given case studies. One workshop lasted two days and was directed to MRes Design students, the other one lasted three hours and was directed to cross-discipline PhD students. The first and longer workshop started with a lecture given by an ethics and computer scientist PhD student, who discussed with the students the contemporary ethical challenges of the digital environment and the effects these have on people's attitudes and behaviours. The lecture was followed by a group activity where students had to map the ethical issues of two case studies: one was a chatbot enduring life after death, the other looked at the interface of an online news platform. The mapping needed to reflect the ethical issues raised from the different parties involved in the case studies. To achieve so students "became" these parties. In day two the two groups developed negotiated ethical principles. The group of students analysing the chatbot case study took the roles of the manufacturer, the designer, the distributer and the consumer. Altogether the group discussed each of these party's perspectives, including any personal interest with the related impact on the actual customer. In this case study students understood a conflict in the customer role, who could equally be the deceased and the message recipient. This paradox impacts the mapping exercise. Through the debate the group agreed on the deceased's responsibility of causing psychological distress to the recipient (Fig.01).

The second group decided to list the values each single role could bring to the project before mapping the ethical issues; the students understood that different cultural and professional backgrounds impact the decision-making process (Fig.03). To overcome this issue students created a "value bank", i.e. a bank that stores the values the group shares. Together with the "value bank" the principle used to develop the news interface was the user's perspective; this was to acknowledge the impact of fake news on readers.
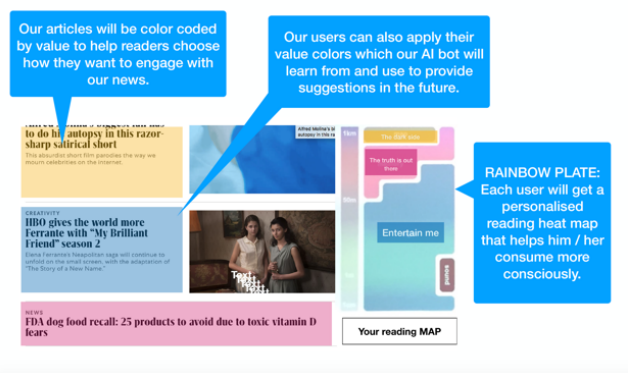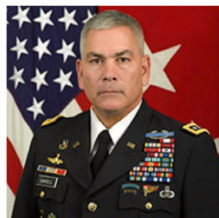


*Fig.01(lx) Chatbot fictional message that illustrates the ethical issue. Fig.02 (rx) Colour coded interface*

To mitigate this the group -i.e. the CEO, the designer, the journalist and the reader - designed an interface ruled by the interactions of the different parties (Fig.04-02). This is operated by a colour coded system that aims at instilling trust and transparency between readers and the editorial board. Colours are the system that gives readers agency and enable mutual trusted relationships.

| CEO | JOURNALIST | WEB DESIGNER | READERS | KEYWORDS | |
|---|---|---|---|---|---|
| Hierarchy | Truth | Unique | Truth | Manage, Profit, Community, Hierarchy, Whole, Balance, Unique, Conscious, | Adaptable, Anchored, Diverse, Floating, Tolerant, Empathy, Equal, Responsibility, Freedom. |
| Diverse | Conscious | Empathy | Freedom | | |
| Profit | Diverse | Whole | Diverse | | |
| | | Balance | | | |

*Fig.03. (lx) Value Bank. Fig.04 (rx) Roles performed by the group*

Stepping into someone's shoes helped both groups develop the first-hand experience that put them in the decision-making psychology. The exercise succeeded in engaging students with criticality and reflection which helped construct the debate and deliver ethics design principles.

In the second shorter workshop doctoral students were asked to divide in smaller groups across the four Schools to analyse, map and define ethical principles of given case studies. The case studies were three in total but the six groups gravitated to one, which was the development of a smart wearable that patients could use to store and consult medical data accessible by the National Health Service (NHS). Doctoral students enjoyed the debate raised by the case study. Even though the selection was uniform across the groups this didn't affect the richness of the debate; discussions amongst groups varied - data storage, marketing, accessibility- and it was driven by the personal research area and expertise. The workshop succeeded in engaging students with ethics through debate, discussion and reflection.

## 3    REFLECTIONS

Looking at the three examples analysed in this paper (EthiCS, Berkley- Delft and RCA) there are common parameters emerging across all cases; these are collaboration of different parties, integration of ethics in existent practices, inter/cross-disciplinary engagement and culture. These parameters have been crucial in determining the success and failure of each example. They all evidenced how teaching ethics is an approach that thrives or fails in relation to the cultural and social environment and the level of engagement. In the RCA workshops culture was used as an exchangeable factor capable of outlining the relations between terms and events, as causes and effects. Culture was then the factor pointing at *difference and differentiation* [3]. This evidences that ethics is a practice that can be taught by focusing on the differential modalities of interactions, which balance creates the environment for debate, exchange and reflection [3]. Under this light there is an important factor to take into consideration, which is the role of history. History determines the values people stand for, as they are embodied in the social and historical backgrounds in which people grew up. One of the RCA workshops outlined this aspect. Students felt necessary to put on the table their values before starting the mapping exercise. Values articulate the underlying historical factor embedded in ethics, which is an aspect that determines the way people look at the future [10]. Ethics are the representation of the past enduring in the future. The news website case study group recognised this historical factor and the impact it plays when ethics is a common good. Students indeed found necessary to design a "value bank" that could support alignment of decisions. Of course in a real scenario this wouldn't be as simple; the collection of values need to account the emotional aspect in any personal interest. Nonetheless assuming that the group can overcome personal interests through negotiation, the act of questioning, which is inherent to ethics, makes the practice an *intersubjective* [10] process: i.e. a subject depends on the other but also recognises the value of the other. This helps build trust against parties. Under this light ethics become a "mediation" process, i.e. a recognition of the *other's claims on me* [10]. When this is acknowledged individuals become keener to listen and negotiate with others. Under these terms ethics can be described as a dynamic process that implies a consistent dialogue and negotiation between individuals (and things) and as a *movement of "getting lost" that opens up the space for the ethical work of mediation which alone generates the solution* [10]. Feedbacks from the doctoral students identified the engaged aspect as the most valuable lesson learnt; indeed, to discuss ethics groups had to design a debate based on what each person believed, even though this was not one of the workshop requirements. The second workshop didn't succeed in designing universal principles. To overcome this aspect, it would have been better to facilitate the activities to guide the discussion with templates.

# 4    CONCLUSIONS

This paper articulated the value of practising and teaching ethics in STEAM environments with the ambition and motivation to develop technological infrastructure that supports social innovation and respects human agency. Two workshops described how collaborative and engaged design research can support the practice of ethics across disciplines and facilitate discussion amongst participants, who are called to take equally responsibility of the decision-making process. The workshops articulated how the practice of ethics can stimulate change through debate and participation. Even though the relationship between human and machine has been not directly investigated and analysed, the workshops enabled a discussion where technology was the subject, but analysed through the human lenses. The impacts of technology in society were the parameters the groups took as reference to focus the attention on the human perspective, which was the real subject of discussions. By moving the human to the foreground, and analyse the case studies from this perspective, it was possible to draw ethical codes designed around people. The faceted, multiform and dynamic nature of ethics proved to be an instrument to question issues, roles and responsibilities by acknowledging the diversity of participants. Ethics can design opportunities to tackle complex issues across sectors and disciplines by being a method of communication across the parties involved in the process. Practicing debate and reflection can help expose issues and challenges and negotiate these with a group of different people. This methodology can lead the way to a new culture of ethics and technology aimed at tackling and addressing complex problems through human cognitive skills. However to succeed in this it is necessary to design this culture. The workshops illustrated in this paper were aimed at drawing guidelines, which need implementation and further studies and research. Further work is indeed required to test this methodology outside academia and with experts working across the life cycle of products, services and infrastructures. The workshops findings demonstrate that rethinking communication and human-machine-human interaction through ethics can redesign hierarchies and organizational charts, as well as the distribution of roles and tasks between people and technology. The workshops also outline the value of these relationships and, in particular, the value of the exchange of information, physical and digital, which is key in any industry. The design of a different socio-technological development requires the engagement of cross-sectors/disciplines communities. They can develop a different decision-making and validation process, which is rounded on those involved directly and indirectly in the process - both machines and humans. With ethics as a practice that can't be codified but needs consistent update, human-machine- human interaction can develop cross-sector dynamic practices that support the human factor with its creativity in tackling issues and challenges.

## REFERENCES

[1] Embeded EthiCS, (2019). in https://embeddedethics.seas.harvard.edu/publications.html [Accessed in March 2019]

[2] Bathaee, Y. (2018). The artificial intelligence black box and the failure of intent and causation. In *Harvard Journal of Law & Technology*, 31(2), 889.

[3] McCormack (2018), "Posthuman Ethics", in Braidotti, R., & Hlavajova, M. (Eds.). (2018). *Posthuman glossary*. Bloomsbury Publishing.

[4] Braidotti, R., & Hlavajova, M. (Eds.). (2018). *Posthuman glossary*. Bloomsbury Publishing.

[5] Grosz, Barbara, David Gray Grant, Kate Vredenburgh, Jeff Behrends, et al. 2018. Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education. Forthcoming - to be published in Communications of the ACM (CACM)

[6] Kissinger, H. A.,"How the Enlightenment Ends", in "*The Atlantic*" Online June 2018, in https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/ [Accessed in March 2019]

[7] Morus, I.R. 2002. Bodies/Machines, Bloomsbury Press

[8] Öhman, C., & Floridi, L. (2017). The political economy of death in the age of information: A critical approach to the digital afterlife industry. Minds and Machines, 27(4), 639-662.

[9] Taebi, B., & Kastenberg, W. E. (2016). Teaching Engineering Ethics to PhD Students: A Berkeley–Delft Initiative. Science and Engineering Ethics, 1-8.

[10] Žižek, S. (2009). The parallax view. MIT Press.