DIGITAL ETHICS CANVAS: A GUIDE FOR ETHICAL RISK ASSESSMENT AND MITIGATION IN THE DIGITAL DOMAIN

C. Hardebolle [1]
Center for Digital Education, Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
https://orcid.org/0000-0001-9933-1413

V. Macko
Information Management Institute, Université de Neuchâtel (UniNE)
Neuchâtel, Switzerland
https://orcid.org/0009-0003-8228-0404

V. Ramachandran
Teaching Support Center, Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
https://orcid.org/0000-0001-5249-2578

A. Holzer
Information Management Institute, Université de Neuchâtel (UniNE)
Neuchâtel, Switzerland
https://orcid.org/0000-0001-7946-1552

P. Jermann
Center for Digital Education, Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
https://orcid.org/0000-0001-9199-2831

ABSTRACT

Ethical concerns in the digital domain are growing with the extremely fast evolution of technology with machine learning techniques and the increasing scale at which software is deployed, potentially affecting our societies globally. It is crucial that engineers evaluate more systematically the impacts their solutions can have on individuals, groups, societies and the environment. Ethical risk analysis is one of the approaches that can help reduce "ethical debt", the unpaid cost generated by ethically problematic technical solutions. However, previous research has identified that novices struggle with the identification of risks and their

---

[1] Corresponding Author: C. Hardebolle, cecile.hardebolle@epfl.ch

mitigation. Our contribution is a visual tool, the Digital Ethics Canvas, specifically designed to help engineers scan digital solutions for a range of ethical risks with six "lenses": beneficence, non-maleficence, privacy, fairness, sustainability and empowerment. In this paper, we present the literature background behind the design of this tool. We also report on preliminary evaluations of the canvas with novices (N=26) and experts (N=16) showing that the tool is perceived as practical and useful, with positive utility judgements from participants.

1 INTRODUCTION

The ethical issues with software, especially Artificial Intelligence (AI)-based software, released to the public are so widespread that some researchers have coined the term "ethical debt" (Petrozzino 2021; Fiesler 2020). Paralleling the notion of technical debt (Knesek 2016), ethical debt represents the unpaid cost generated by ethically problematic software and borne by individuals, communities, and society in general. However, there is an important distinction between the two types of debt: while technical debt is typically the result of deliberate choices guided by specific imperatives (e.g. time to market), ethical debt mostly arises from unidentified ethical risks (Fiesler 2020; Petrozzino 2021).

Engineering education has a responsibility to address this situation. Isaac et al. (2022) have shown that novice software engineers tend to neglect ethical concerns in their design. Griffin et al. (2023) report that experienced software engineers do not necessarily identify that technical decisions they routinely make in their professional activities have ethical implications. Whether they will be users, integrators, designers or developers of software systems, the engineers we train need to develop strategies for a) systematically identifying and assessing ethical risks associated with digital solutions[2] and b) identifying possible mitigation options to the ethical issues they identify.

In this paper, we present a visual guide called the "Digital Ethics Canvas" designed to help engineering students work through ethical risks specifically related to digital solutions. We first review previous work on ethics-related canvases before discussing the foundations for the ethical framework underlying our canvas. We then present preliminary evaluation results.

2 BACKGROUND

A "canvas" is a visual tool designed to guide people through the process of using a methodology or framework. Canvases are increasingly used in engineering education (Tranquillo, Kline, and Hixson 2016), for instance to support specific engineering tasks (Ruf and Back 2015) or to support education activities in an engineering context (Ammersdörfer et al. 2022). The canvas that we propose has two specificities: a) it focuses on the analysis of the risks generated by a digital solution under design, development, or use and b) it is built on six ethical "lenses" that guide risk assessment and mitigation. In the following, we position our approach compared to existing work on these two aspects.

2.1 Risk analysis

Risk analysis is an important aspect of engineering work and encompasses two types of risks: risks to the product/service being engineered (i.e. hazards that could make the

---

[2] Throughout this paper, we use the term "digital solution" to refer to technical systems that involve software and the infrastructure needed to run it. We chose this term to encompass a wide range of software-related technologies, such as IA, Internet of Things, Big Data, social networks or web apps.

engineering project fail) and risks generated by that product/service (i.e. adverse effects that the product/service could have on individuals, groups, societies and the environment). While the former are important from a project and business management point of view, the latter are at the core of responsible engineering and the focus of our work. Our goal is to develop the ability of engineers to identify the specific risks generated by digital solutions to others, society and the environment (also called ethical sensitivity).

Vallor (2018) proposes "ethical risk sweeping" as a tool for avoiding "ethical negligence". She argues that ethical risks analysis should be a standard engineering protocol in the same way as cybersecurity penetration testing, and repeated at all phases in the engineering process, from the initial product proposal to the quality assurance stage. However, her proposal does not make explicit how engineers can analyze these ethical risks in practice.

Carlson et al. (2018) have studied how students analyze risks in the context of projects involving real-world problem solving. They found that students struggled with three aspects of risk analysis: identifying risks, setting priorities and working on mitigation (Carlson et al. 2018). They have proposed two canvases, a Design Canvas and an Iteration Plan, to support students in analyzing risks during iterative project planning: the Design Canvas is used to identify risks in the problem space and the Iteration Plan to prioritize and set mitigation goals. However, they define risk as "the probability that the design project fails to make impact", which does not include risks generated by the impact the design project will make.

Among other methods frequently used in strategic analysis, the SWOT matrix (Strengths, Weaknesses, Opportunities, Threats) includes a risk analysis component within its "Threats" section (Weihrich 1982). However, its main drawback from our perspective is that it also considers only risks to the project. A similar drawback is found in other canvases designed specifically for risk analysis by adapting the Business Model Canvas (BMC) from Osterwalder et al. (2010), such as in (Borbinha, Nadali, and Proença 2015) or (Kuru and Artan 2020).

Taking a different approach, Reijers et al. do not focus on risks per se but on "how a technology might bring about ethical impacts for different stakeholders" (Reijers et al. 2018). Designed for practitioners who do not necessarily have an ethics background, their "Ethics Canvas" (which is also adapted from the BMC) implements a four-phase process, from stakeholder and impact analysis to mitigation design. Two clear strengths of the Ethics Canvas are its focus on risks generated by a product/service and its domain-agnostic approach, which makes it suitable for a wide range of disciplines and applications. On the other hand, it might be difficult for novice engineers to think about certain ethical concerns in their analysis in the absence of specific guidance. For instance, while privacy and fairness issues seem to gain increasing visibility in software engineering, other concerns such as sustainability and environmental impacts seem to be less frequently addressed (Isaac et al. 2022). In the next section, we review different approaches that try to tackle this issue.

2.2 Ethical lenses

Value-oriented methodologies such as Value-Sensitive Design (Friedman, Kahn, and Borning 2002) typically approach the range of ethical concerns by having stakeholders identify explicitly the "human values" that the product/service should align with. Value-based approaches are getting a lot of traction in engineering, with some researchers proposing Value-Based Engineering as an "ethics by design" approach (Spiekermann and Winkler 2020). The focus on values in these approaches can be seen both as a strength and a weakness: on one hand, the contextual nature of values makes these approach flexible and adoptable to a broad range of contexts, but on the other hand, appropriately defining the

values at stake and frame them so that they mean the same thing to all parties can be challenging (Friedman et al. 2021). The very concept of value has actually been deemed unclear and insufficiently defined (Manders-Huits 2011).

Some authors argue for a more normative approach based on predefined ethical principles (Manders-Huits 2011). Cardia et al. (2017) for instance, propose a canvas based on the four humanitarian principles (humanity, neutrality, impartiality and independence) to assess the use of digital technology for humanitarian action. The four bioethics principles (beneficence, non-maleficence, autonomy, and justice) are often used as an ethical framework for evaluating engineering solutions in healthcare contexts, such as in (Cawthorne and Robbins-van Wynsberghe 2020). By definition, principle-based approaches are possible only when there is an overall agreement on the set of ethical principles to use. This is the case for the humanitarian domain (Council of the EU, European Parliament, and European Commission 2008) and for the medical domain with the largely adopted principles of biomedical ethics (Beauchamp and Childress 1979). As we will discuss in the next section, this is not (yet) the case for the digital domain. In addition, the engineers we train will work in a variety of contexts with varying sets of values (e.g. healthcare, social media). This is why we adopt instead the notion of ethical "lenses" as proposed by Isaac et al. (2022), which represent multiple ethical perspectives for analyzing risks.

The five "lenses" from Isaac et al. (2022) stem from the three human-centered design criteria originally proposed by the design company IDEO (IDEO 2000) - feasibility, desirability, viability, that the authors extended with sustainability, privacy and accessibility. The idea of guiding analysis with several criteria is also found in the PEST/PESTLE framework (Political, Economic, Social, Technological / Legal, Environmental). This analysis grid is meant for "scanning" macro-environmental factors in business development (Aguilar 1967) and is often used in combination with SWOT. Other authors have instead reinterpreted existing canvases in light of such criteria. For instance, Gillet et al. (2022) propose two reinterpretations of the Value Proposition Canvas (Osterwalder et al. 2014), focused respectively on sustainability and transparency. In contrast to this latter approach and with the goal to help engineers "scan" risks from multiple ethical perspectives, we propose a single canvas that implements several ethical lenses. In the following section, we discuss the ethical lenses we chose in light of the existing literature on ethics in the digital domain.

2.3 Digital-specific ethical lenses

With the increasing visibility of ethical issues with software, researchers and practitioners have attempted to clarify ethical guidelines for the digital domain. A significant number of proposals stem from the Big Data and Artificial Intelligence domains. In their "Data, responsibly" proposal, Stoyanovich, Abiteboul, and Miklau (2017) recommend fairness, diversity, transparency, equality and data-protection as the foundations for responsible data science. One years later, Ballantyne (2018) argues that "there is no one-size-fits-all framework for how to ethically manage your data" and suggests seven ethical values for "making informed, explicit, and justifiable trade-offs, rather than following a set of prescribed rules": social value, harm minimization, control (autonomy), justice, trustworthiness, transparency and accountability. Howe and Elenberg (2020), take a medical research stance and suggest autonomy, equity and privacy as the ethical concepts most challenged by the use of big data in health. Interestingly, the issue of sustainability and environmental impact is mostly absent from these proposals.

In the domain of AI, Jobin, Ienca and Vayena (2019) analyzed 84 documents in the context of the "ethical AI debate" to identify whether a global consensus was emerging. They identified that 88% of the documents had been published after 2016 and conclude that "No single ethical principle appeared to be common to the entire corpus of documents, although there is an emerging convergence around the following principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy.". Loi, Heitz and Christen (2020) extended this work with a focus on the procedures recommended in these AI ethics guidelines. As a result, they propose a framework with seven principles: beneficence, non-maleficence, autonomy, justice, control, transparency, accountability. Ryan and Stahl (2020) also extended the work from Jobin et al. but with the goal of providing the most comprehensive list of ethical principles: they present the ethical principles found in 91 guidelines organized into 11 themes. It is the only contribution we found that included sustainability and environmental impact, reflecting an overall lack of attention to a pressing issue to which digital solutions are actually no stranger (Bender et al. 2021). Other authors take a radically different approach and put forward the human rights framework as a cross-cultural and globally agreed framework for responsible AI (Prabhakaran et al. 2022).

With this short review we want to highlight the current lack of consensus on the ethical principles that should guide a responsible approach to software. It is important to note that this landscape is moving extremely fast and is influenced, of course, by the crucial work done on software and AI regulation worldwide. A flagship of this work is probably the "Artificial Intelligence Act" from the European Commission, which follows a risk-based approach to classify AI-based systems in terms of impacts on safety, security and fundamental rights.

In terms of canvas-based approaches, we found one digital-specific ethics canvas: the Technology Impact Cycle Tool (TICT) (Fontys University 2021). Focused on reflection, it uses questions organized in "scans" of progressive scope with 10 different categories: impact on society, human values, privacy, inclusivity, transparency, bad actors, sustainability, data, stakeholders and futuring. While we found the organization in progressive scopes helpful, we thought that the tool had too many categories and was mixing design process aspects (e.g. stakeholder analysis) with ethical lenses (e.g. privacy, sustainability). We also argue that, while reflection is certainly important in responsible design, an analytical approach of the risks generated by a solution is essential to reducing ethical debt.

3 THE DIGITAL ETHICS CANVAS

As our review highlighted, very few options exist to help engineers identify and mitigate the range of ethical risks generated by a digital solution under design, development or use. Our contribution is a canvas, depicted in Figure 1, that helps engineers to scan the risks generated by a solution with six digital-specific ethical lenses: beneficence, non-maleficence, privacy, fairness, sustainability and empowerment.

Following an incremental process and taking inspiration from the bioethics principles in particular, we have integrated ethical lenses progressively into our canvas. Our "beneficence" lens is positively oriented, for documenting the expected benefits of the solution. The "non-maleficence" lens is meant to capture safety and security issues, as suggested by Ryan and Stahl (2020). Our "empowerment" lens reflects the autonomy principle but with a larger scope to encompass issues of transparency, explainability, trust and user agency. Our review showed that "justice" and "fairness" are often grouped together (see previous section) but we chose to use "fairness" as a less normative concept which is more frequently used in relation to AI-based solutions. Finally, we add a "privacy" lens to capture risks with regards to the use

of data and "sustainability" to include risks related to environmental impacts and labor exploitation. An important factor in our choice was to limit the number of lenses not to overwhelm novices, while being general enough to capture a range of ethical risks. We were also careful not to be too Big Data- or AI-specific and made sure our canvas can be applied to other types of digital solutions. For scaffolding purposes, we included questions in the canvas for the different lenses to help our users surface elements in the digital solution that are likely to give rise to risks, rather than providing them with definitions (see Figure 1).

Context	Beneficence	Non-maleficence		Solution
□ In which context is the solution evaluated?	□ What are the expected benefits of the solution in this context?	**Risks**	**Mitigation**	□ What are the characteristics of the solution under evaluation?
		□ Can the solution be used in harmful ways, in particular with regards to vulnerable populations? □ What kind of impacts can errors from the solution have? □ What type of protections does the solution have against attacks?		
	Privacy	**Fairness**		
	Risks / **Mitigation**	**Risks** / **Mitigation**		
	□ What data does the solution collect? □ Is it collecting personal or sensitive data? □ Who has access to the collected data? □ How is the collected data protected?	□ How accessible is the solution? □ What kinds of biases may affect the results? □ Can the outcomes of the solution be different for different users or groups?		
	Sustainability	**Empowerment**		
	Risks / **Mitigation**	**Risks** / **Mitigation**		
	□ What is the carbon footprint of the solution? □ What types of resources does it consume (e.g. water) and produce (e.g. waste)? □ What type of human labor is involved?	□ Can users understand how the solution works and what its limits are? □ Are users able to make choices (e.g. consent, settings) in their use of the solution and how? □ How does the solution affect user autonomy and agency?		

Figure 1: The Digital Ethics Canvas. The left and right columns are used to map out factual information about the digital solution and the context, whereas the central part is used for evaluating the benefits, risks and mitigation options for the solution using our ethical lenses.

With one benefit-oriented lens and five risk-oriented lenses, our canvas supports users in benefit-risk analysis, a methodology which is widely used in public health - see for instance (European Medicines Agency 2018). Benefit-risk analysis is part of a broader family of Multi-Criteria Decision-Making methods (Zionts 1979), typically used in the case of multiple conflicting objectives, as is generally the case with ethical decisions. This type of analysis requires collecting information about the problem space and context, which is why our canvas includes sections to map out factual information about the digital solution and its context of use. For each of the risks, mitigation strategies can be described, as these can weigh in the analysis. Depending on whether the canvas is used at design/development or at use time, mitigation strategies can either be artifact-oriented or context-oriented. That is, the strategies could include changing the technological artifact (e.g., avoid collecting personal data that is not needed to reduce a privacy risk) or changing the usage context (e.g., ask users to provide a nickname rather than their actual name).

4 EVALUATION

4.1 Methods

We have developed our canvas incrementally, testing our ethical lenses and our approach with different types of audiences and applications. In this paper, we report the results of two small-scale evaluations conducted in the spring semester 2023. The first evaluation collected the views of novices and took place in a master's course on Information Systems Design with 26 students of various backgrounds (15 women, 11 men). The evaluation was part of a three-hour session dedicated to responsible design. After an introduction dedicated to fairness in AI, we facilitated an interactive presentation of the canvas and its ethical lenses, then students applied the canvas on a case study, followed by a class discussion. The second evaluation was a part of a 90-minute workshop on the ethics of using generative AI for education, for experts in the fields of ethics, engineering, and education. The experts had various backgrounds and levels of seniority (N=16, 11 women, 5 men). We introduced our ethical lenses one after the other with inputs from research on generative AI and time for analyzing the corresponding risks in a given scenario (e.g., a teacher using generative AI to generate deepfakes instead of recorded lectures).

At the end of each session, we asked participants to fill out a survey in which we considered both affective reactions and utility judgment measures (Alliger et al. 1997). We captured participants' perceptions about the canvas with the AttrakDiff questionnaire (Hassenzahl, Burmester, and Koller 2003), which has 10 items with pairs of opposite adjectives and a scale from 1 to 7, every other item being reverse-scored. In addition, participants were asked "How would you describe the canvas to your friends?". The second part of the survey asked participants their perceptions about several aspects of the workshop facilitation, using a 4-point likert scale. To assess the perceived utility of the canvas, we asked participants if they thought what they had learned in the session would be useful to them later, if they were likely to apply what they learned in other contexts and if they wanted to have access to the canvas for further use (also on a 4-point likert scale). For novices, learning outcomes were also evaluated in a mid-term exam question the following week.

4.2 Results: perceptions about the canvas

The results of the AttrakDiff items are shown in Table 1. All the adjective pairs are presented with the negative adjective on the left (valued 1) and the positive on the right (valued 7), taking reverse-scoring into account. For each question, we indicate the mean score and standard deviation for both the novices and the experts. The star notation indicates when the mean score is significantly different from neutral (4) as assessed with a single-sample t-test.

The most statistically significant measures indicate that both novices and experts found the canvas 'good' (M = 5.92 for novices and M = 6.44 for experts respectively), 'practical' (M = 5.35 and M = 6.06) and 'useful' (M = 5.84 and M = 6.50). One more measure is statistically significant for the experts suggesting they also found the canvas 'stylish' (M = 5.81). All other mean scores are above the neutral value (4 of 7) and a majority higher than 5. These results indicate that the canvas was perceived positively and was found to possess pragmatic qualities by both the novice and the expert audiences, a valuable attribute for an education tool for engineers. Overall, the experts report a more positive perception of the canvas than novices. While this is probably to be expected, it also indicates for us an opportunity to further tailor the canvas for novice users, for instance by developing the educational resources around our canvas (e.g. an accompanying quick start guide). The only scale where

experts rated the canvas lower than novices is the "complicated - simple" scale. We relate this to the fact that experts had much less time to practice with the canvas than novices.

Table 1: Results of the AttrakDiff questionnaire, with differences from the neutral response (4) tested with single-sample t-tests (*** for p < .001, ** for p < .01 and * for p <.05).

	Negative pole (1)	Positive pole (7)		Novices (N=26)	Experts (N=16)
Attractiveness (ATT)	Bad	Good	Mean (SD)	**5.92***** (1.12)	**6.44***** (0.81)
	Ugly	Attractive	Mean (SD)	4.5 (1.57)	5.56 (1.21)
Hedonic Quality - Identity (HQ-I)	Unimaginative	Imaginative	Mean (SD)	5.35 (1.26)	6.00 (0.89)
	Dull	Captivating	Mean (SD)	5.23 (1.61)	6.25 (0.58)
Hedonic Quality - Stimulation (HQ-S)	Tacky	Stylish	Mean (SD)	5.12 (1.48)	**5.81**** (1.05)
	Cheap	Premium	Mean (SD)	4.92 (1.38)	5.81 (1.11)
Pragmatic Quality (PG)	Confusing	Clearly structured	Mean (SD)	5.73 (1.22)	5.94 (1.29)
	Complicated	Simple	Mean (SD)	5 (1.10)	4.31 (1.54)
	Impractical	Practical	Mean (SD)	**5.35*** (1.44)	**6.06***** (0.85)
	Useless	Useful	Mean (SD)	**5.84***** (1.25)	**6.50***** (0.52)

Most of the novices provided a description of the canvas, with mostly positive responses such as: "A simple model to assess complex 'blurry' topics", "Tool to help you remember/realize the different aspects of a project that you don't necessarily think of intuitively." and "As a good way to break down and analyze the important aspects of different phenomena". A few responses were mixed, such as: "A good tool but could be simpler." and "Complicated but interesting". Only a few experts provided a textual description of the canvas, such as "A useful thinking tool.", "Matrix for structured evaluation" or "As a useful tool to consider in teaching witch's of AI". It is interesting to note how a number of novices contrast the practicality of the canvas with the complexity and non-intuitive aspects of technology ethics, which we take as a positive indicator of the value of the canvas for educational purposes. The more mixed descriptions by novices further encourage us to refine the instructional design of our canvas and our introduction session.

4.3 Results: quality of the session and utility judgements

The perceptions of novices and experts about the facilitation of the sessions are illustrated in questions 1 to 4 of Figure 2. The majority of participants agreed or strongly agreed with all the positive statements. Notably, 95.9% of novices and 100% of experts found the session good. Experts were a bit less positive than novices on the time allowed for questions, which we relate to the relative short duration of the workshop compared to the course.

Utility judgements of the participants are represented in questions 5 to 7 of Figure 2. Participant's evaluation of the usefulness of the session is positive with 75% of novices and 87.6% of experts agreeing or strongly agreeing. A high proportion of participants (66.6% of

novices and 81.3% of experts) also reported they were likely to apply what they had learned into other contexts, which is a positive indicator of both learning and transfer (Alliger et al. 1997). Finally, the majority of respondents (62.5% of novices and 75% of experts) indicated they would like to have access to the canvas for other tasks. While this evaluation is promising, we observe that novices are overall less positive in their utility judgment than experts. This is to be contrasted with the results in terms of learning outcomes, where novices scored an average grade of 74%. Their less positive utility judgments might be due to the fact that, at the time of the session, they had not yet started to work on their course project and might not have directly seen how to apply the canvas in a concrete context. If that explanation proves to be correct, it would underline the importance of combining such an ethical canvas session with a concrete real-life project.
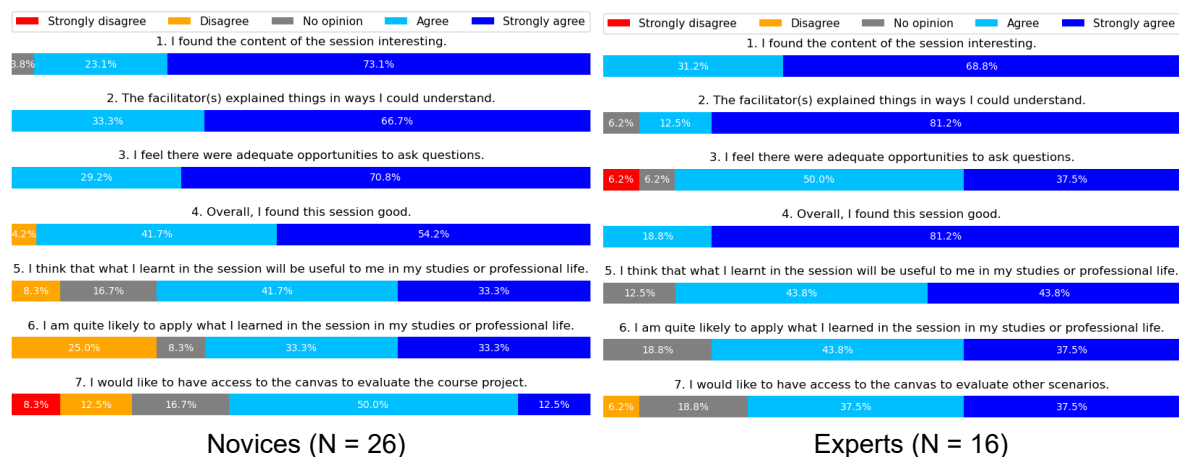

Figure 2: Results of the questionnaire (4-item Likert scale) on the perceptions about the session (questions 1 to 4) and utility judgements on the canvas (questions 5 to 7).

5   SUMMARY AND ACKNOWLEDGEMENTS

In this paper, we address the issue of ethical debt by proposing a canvas for engineers to analyze more systematically the ethical risks with a digital solution at design, development or use time. Our canvas includes six ethical lenses (beneficence, non-maleficence, privacy, fairness, sustainability and empowerment) that are specific to the digital domain and implement a benefit-risk analysis framework. We presented an overview of the literature behind our design choices as well as a preliminary evaluation of our canvas by engineering ethics novices and experts. The results proved positive, canvas being perceived as practical and useful by participants, which is promising for use in engineering education. We plan to further refine our instructional resources around the canvas to provide users with more scaffolding, and to further evaluate how this canvas can be used in various settings.

BIBLIOGRAPHY

Aguilar, F. J. 1967. *Scanning the Business Environment*. Johannesburg: Macmillan.
Alliger, George M., Scott I. Tannenbaum, Winston Bennett Jr, Holly Traver, and Allison Shotland. 1997. "A Meta-Analysis of the Relations Among Training Criteria." *Personnel Psychology* 50 (2): 341–58. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x.

Ammersdörfer, Theresa, Darien Tartler, Simone Kauffeld, and David Inkerman. 2022. "Reflection Canvas – An Approach to Structure Reflection Activities in Engineering Design." In *DS 118: Proceedings of NordDesign 2022, Copenhagen, Denmark, 16th - 18th August 2022*, 1–12. https://doi.org/10.35199/NORDDESIGN2022.29.

Ballantyne, Angela. 2018. "Where Is the Human in the Data? A Guide to Ethical Data Use." *GigaScience* 7 (7): giy076. https://doi.org/10.1093/gigascience/giy076.

Beauchamp, Tom L., and James F. Childress. 1979. *Principles of Biomedical Ethics*. New York: Oxford University Press.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? □." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Borbinha, José, Ahmad Nadali, and Diogo Proença. 2015. "A Pragmatic Risk Assessment Method Supported by the Business Model Canvas:" In *Proceedings of the Fifth International Symposium on Business Modeling and Software Design*, 156–62. Milan, Italy: SCITEPRESS - Science and and Technology Publications. https://doi.org/10.5220/0005886501560162.

Cardia, Isabelle Vonèche, Adrian Holzer, Ying Xu, Carleen Maitland, and Denis Gillet. 2017. "Towards a Principled Approach to Humanitarian Information and Communication Technology." In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, 1–5. ICTD '17. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3136560.3136588.

Carlson, Spencer Evan, Leesha V. Maliakal, Daniel Rees Lewis, Jamie Gorson, Elizabeth Gerber, and Matthew Easterday. 2018. "Defining and Assessing Risk Analysis: The Key to Strategic Iteration in Real-World Problem Solving," July. https://repository.isls.org//handle/1/775.

Cawthorne, Dylan, and Aimee Robbins-van Wynsberghe. 2020. "An Ethical Framework for the Design, Development, Implementation, and Assessment of Drones Used in Public Healthcare." *Science and Engineering Ethics* 26 (5): 2867–91. https://doi.org/10.1007/s11948-020-00233-1.

Council of the EU, European Parliament, and European Commission. 2008. "The European Consensus On Humanitarian Aid." *Official Journal of the European Union* 51 (2008/C 025/01). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A42008X0130%2801%29.

European Medicines Agency. 2018. "Benefit-Risk Methodology." Text. European Medicines Agency. September 17, 2018. https://www.ema.europa.eu/en/about-us/what-we-do/regulatory-science-research/benefit-risk-methodology.

Fiesler, Casey. 2020. "Ethical Tech Starts With Addressing Ethical Debt." *Wired*, September 16, 2020. https://www.wired.com/story/opinion-ethical-tech-starts-with-addressing-ethical-debt/.

Fontys University. 2021. "Technology Impact Cycle Tool." Technology Impact Cycle Tool. 2021. https://www.tict.io/.

Friedman, Batya, Maaike Harbers, David G. Hendry, Jeroen van den Hoven, Catholijn Jonker, and Nick Logler. 2021. "Eight Grand Challenges for Value Sensitive Design from the 2016 Lorentz Workshop." *Ethics and Information Technology* 23 (1): 5–16. https://doi.org/10.1007/s10676-021-09586-y.

Friedman, Batya, Peter H Kahn, and Alan Borning. 2002. "Value Sensitive Design: Theory and Methods." *University of Washington Technical Report*, 8.

Gillet, Denis, Isabelle Vonèche Cardia, and Jérémy Alain La Scala, eds. 2022. "Introducing Alternative Value Proposition Canvases for Collaborative and Blended Design Thinking Activities in Science and Engineering Education." *[Proceedings of TALE 2022]*.

Griffin, Tricia A., Brian Patrick Green, and Jos V. M. Welie. 2023. "The Ethical Agency of AI Developers." *AI and Ethics*, January. https://doi.org/10.1007/s43681-022-00256-3.

Hassenzahl, Marc, Michael Burmester, and Franz Koller. 2003. "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität." In *Mensch & Computer 2003: Interaktion in Bewegung*, edited by Gerd Szwillus and Jürgen Ziegler, 187–96. Berichte des German Chapter of the ACM. Wiesbaden: Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-80058-9_19.

Howe, Edmund G., and Falicia Elenberg. 2020. "Ethical Challenges Posed by Big Data." *Innovations in Clinical Neuroscience* 17 (10–12): 24–30.

IDEO. 2000. "How to Prototype a New Business." IDEO U. 2000. https://www.ideou.com/blogs/inspiration/how-to-prototype-a-new-business.

Isaac, Siara, Aditi Kothiyal, Pier Luca Borsò, and Bryan Ford. 2022. "Someone Else's Problem – Sustainability and Ethicality Are Peripheral to Students' Software Design." *International Journal of Engineering Education*, 17.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

Knesek, Doug. 2016. "Averting a 'Technical Debt' Crisis (Part 1)." LinkedIn. 2016. https://www.linkedin.com/pulse/averting-technical-debt-crisis-part-1-doug-knesek/.

Kuru, Kadir, and Deniz Artan. 2020. "A Canvas Model for Risk Assessment and Performance Estimation in Public–Private Partnerships." *International Journal of Construction Management* 20 (6): 704–19. https://doi.org/10.1080/15623599.2020.1763898.

Loi, Michele, Christoph Heitz, and Markus Christen. 2020. "A Comparative Assessment and Synthesis of Twenty Ethics Codes on AI and Big Data." In *2020 7th Swiss Conference on Data Science (SDS)*, 41–46. Luzern, Switzerland: IEEE. https://doi.org/10.1109/SDS49233.2020.00015.

Manders-Huits, Noëmi. 2011. "What Values in Design? The Challenge of Incorporating Moral Values into Design." *Science and Engineering Ethics* 17 (2): 271–87. https://doi.org/10.1007/s11948-010-9198-2.

Osterwalder, Alexander, Yves Pigneur, Gregory Bernarda, and Alan Smith. 2014. *Value Proposition Design: How to Create Products and Services Customers Want*. Strategyzer Series. Hoboken: John Wiley & Sons.

Osterwalder, Alexander, Yves Pigneur, and Tim Clark. 2010. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken, NJ: Wiley.

Petrozzino, Catherine. 2021. "Who Pays for Ethical Debt in AI?" *AI and Ethics* 1 (3): 205–8. https://doi.org/10.1007/s43681-020-00030-3.

Prabhakaran, Vinodkumar, Margaret Mitchell, Timnit Gebru, and Iason Gabriel. 2022. "A Human Rights-Based Approach to Responsible AI." arXiv. https://doi.org/10.48550/arXiv.2210.02667.

Reijers, Wessel, Kevin Koidl, David Lewis, Harshvardhan J. Pandit, and Bert Gordijn. 2018. "Discussing Ethical Impacts in Research and Innovation: The Ethics Canvas." In *This Changes Everything – ICT and Climate Change: What Can We Do?*, edited by David Kreps, Charles Ess, Louise Leenen, and Kai Kimppa, 299–313. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-99605-9_23.

Ruf, Christian, and Andrea Back. 2015. "How Can We Design Products, Services, and Software That Reflect the Needs of Our Stakeholders? Towards a Canvas for Successful Requirements Engineering." In *New Horizons in Design Science: Broadening the Research Agenda*, edited by Brian Donnellan, Markus Helfert, Jim Kenneally, Debra VanderMeer, Marcus Rothenberger, and Robert Winter, 455–62. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-18714-3_38.

Ryan, Mark, and Bernd Carsten Stahl. 2020. "Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications." *Journal of Information, Communication and Ethics in Society* 19 (1): 61–86. https://doi.org/10.1108/JICES-12-2019-0138.

Spiekermann, Sarah, and Till Winkler. 2020. "Value-Based Engineering for Ethics by

Design." arXiv. https://doi.org/10.48550/arXiv.2004.13676.

Stoyanovich, Julia, Bill Howe, Serge Abiteboul, H V Jagadish, and Gerome Miklau. 2017. "Data, Responsibly." 2017. https://dataresponsibly.github.io/.

Tranquillo, Joe, William Kline, and Cory Hixson. 2016. "Making Sense of Canvas Tools: Analysis and Comparison of Popular Canvases." *Henry M. Rowan College of Engineering Faculty Scholarship*, June. https://doi.org/10.18260/p.26211.

Vallor, Shannon. 2018. "An Ethical Toolkit for Engineering/Design Practice." Markkula Center for Applied Ethics. https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/.

Weihrich, Heinz. 1982. "The TOWS Matrix—A Tool for Situational Analysis." *Long Range Planning* 15 (2): 54–66. https://doi.org/10.1016/0024-6301(82)90120-0.

Zionts, Stanley. 1979. "MCDM—If Not a Roman Numeral, Then What?" *Interfaces* 9 (4): 94–101. https://doi.org/10.1287/inte.9.4.94.