From Value-lists to Value-based Engineering with IEEE 7000™

Sarah Spiekermann
WU Institute for IS & Society
Vienna University of Economics and Business
Vienna, Austria
mis-sek@wu.ac.at

***Abstract*—Digital ethics is currently being discussed worldwide as a necessity to create more reliable IT systems. This discussion, fueled by the fear of uncontrollable general artificial intelligence (AI) and by ethical dilemmas of existing systems, has moved many institutions and scientists to demand value principles that should guide the development of future IT systems. These usually include the demand for privacy, security, transparency, fairness, etc. This article shows why working through lists of values is insufficient for good or ethically aligned design. It will be shown how a truly ethical 'Value-based Engineering' (VbE) would have to look like instead, so that technical product innovation as a whole is put on better (more ethical) feet. VbE is a process-driven, holistic approach to system engineering which initially drew from the ideas of Value Sensitive Design and Ethical Computing. From 2016-2021 VbE was further fleshed out in the IEEE 7000™ standardization project*.**

Keywords—Value-based Engineering, Value Sensitive Design, Ethics, Ethical Engineering, Machine Ethics, Privacy

I. Introduction

In the years since around 2015, there has been a steady increase in awareness of the need to design technology more ethically. Major accidents like the crashes of two Boing 737Max as well as the Volkswagen scandal contributed to the questioning of ethical practices in classic engineering departments. In 2016, the British voted for Brexit and US citizens for Donald Trump. Analyses from the field of investigative journalism later showed that both political decisions were manipulated via social networks; a manipulation process that, on closer inspection, was only possible because the technical architecture of the social platforms, the APIs, the user authentication processes as well as the systematic profiling of users allowed this. At the same time, there are increasing reports of misguided AI systems misleading judges into false prejudices ('biases') or attributing false qualifications to university applicants. Against this background, the IT industry and its customers are becoming aware that more forward-looking, responsible and ethical planning of IT systems is required. Technical values such as privacy, security, transparency, accountability, control, etc. are higher on the engineering agenda than ever before. So high, in fact, that Gartner analysts have begun to highlight technology trends with value predicates; such as '*responsible* AI', '*explainable* AI' or '*private* 5 G'.

But how is this hype about 'more ethical' technology received in practice and how is it absorbed by regular corporate processes? A common approach to the topic of ethics in the everyday practice of system development still seems rather passive. Apart from an occasional exchange with the Corporate Social Responsibility (CSR) department, people in technical business units often tend to only deal with ethics when there is a problem that cannot be ignored; for example, when a corporate audit identifies security gaps or when the EU's General Data Protection Regulation demands steps towards more user privacy. Companies often only become active on ethical design questions when fires have to be put out, when a new law has to be complied with or when values have already been violated so massively that penalties, bad press and customer complaints are on the cards.

That said, the situation might be on the brink of change. More often, companies now publicly commit to value principles. IBM's principles of "Every Day Ethics for AI" or Microsoft's "Responsible AI" list of values are examples. Ann Jobin et al. recently [1] presented an analysis of 84 of such value lists in the journal Nature Machine Intelligence. The hope seems to be that through public statements, guidelines, and even commitments to certain human and social values, there will be an increased awareness of ethics in systems development.

Indeed, *if* the five most accepted values identified by Jobin et al. – that is transparency, fairness, non-maleficence, accountability and privacy- were regularly taught, fully understood, and rigorously implemented as a kind of 'hygiene factors' in all IT systems; *if* they were included in all system engineering textbooks and regularly recognized by startups, then we would certainly witness a major step toward more people-friendly and social digitization. But does the respect of such principle lists allow for saying that a system is really "ethical"? How far does compliance with such lists of value principles take us?

This article discusses the use of value principle lists as an insufficient answer to provide for ethical system design and takes up the sharp criticism recently provided by Brent Mittelstadt, which is that "the truly difficult part of ethics—actually translating normative theories, concepts and values into 'good' practices … —is kicked down the road like the proverbial can." (p.6 in [2]). In response to this criticism, I will show how Value-based Engineering with IEEE 7000™ can address this issue of ethical engineering practice going far beyond value-list compliance.

*This article presents inter alia guidance for ethical engineering given in the forthcoming IEEE 7000™standard. However, this article solely represents the views of the author and does not necessarily represent a position of either the IEEE P7000 Working Group, IEEE or the IEEE Standards Association. The official link to the IEEE P7000 is: https://sagroups.ieee.org/7000/.

II. VALUE LISTS OVERLOOK THE ETHICALLY RELEVANT

In order to grasp what can be understood by ethical design consider the following example of how the value design of voice assistants can be completely different depending on the culture in which they are developed: In 2017, when a user said “I am so sad”, a U.S. Alexa device replied "I wish I had arms to cuddle you". The Russian counterpart Yandex, on the other hand, replied to the same statement "No one said life is a fun event" [3]. Reading these two completely different answers raises the question: Have the developers of the dialog systems actually consciously thought about the ethical implications of the answers given by their system? Have they considered that these express distinct values and that - depending on the market success and diffusion of the system - these could significantly influence users’ attitude to life, for instance children growing up with the system? In fact: Which answer would actually be the more correct one? The American or the Russian one? Thinking about and debating this last question goes to the heart of an ethical discussion on a technical thing: "How should I act?" [4] “I”, that is the engineer of the dialogue system. How do I give the voice assistant the 'right', 'good' or 'wise' dialogue? The answer to this question is a function of the values one wants to pass on with a system.

In the given example, it is up for discussion whether the voice assistant should rather promote the virtue of mental toughness, personality robustness and discipline, or focus on feeling good, conveying closeness and compassion. The decision taken has an important value-ethical impact on society once the respective voice assistants is used at massive scale. It should therefore be answered by innovation teams with a great sense of responsibility; a responsibility that takes time and a degree of work autonomy often not provided these days to system developers [5]. "Value-based Engineering" as embedded in the IEEE 7000™ standard [6] is intended to provide that time and autonomy. It foresees processes to help system development teams and managers to officially engage in value reflection. It involves anticipating the fine-grained value structure of future systems and a deliberate thinking through them in the interest of stakeholders. The identified values are then prioritized and translated into system requirements with the help of so-called "Ethical Value Requirements" (EVRs).

However, the voice assistant example given here implies the recognition of values not included in any of the globally accepted value-lists; not even when one considers the more extended list provided by Jobin et al. [1]. This discrepancy between that which is ethically relevant to a 'system of interest' (SOI) in a context and what is globally recognized or listed as essential is significant. Although public commitment to lists of technical hygiene values is tremendously important it is not sufficient. The broad range of socially relevant values that weave through our interaction with IT systems are much richer and complex than what any preconfigure list can capture.

Lists of value principles are not only too short and overlook relevant, context-driven issues, they also 'prime' the discussion of the true ethical challenges of an SOI. Priming in such a way that the actually relevant value-ethical aspects of a system risk to be neglected. For example, a group of innovation management students at the Vienna University of Economics and Business were invited to think through a shopfloor location-tracking system. The goal of this system was to support senior citizens‘ fashion shopping. One group identified the value of customer privacy as the most important ethical challenge of the tracking system. Prior to their value analysis they had been given a list of values that included privacy. In contrast, a control group of students did not use such a pre-configured value list but instead visited the fashion store physically and talked with the elderly customers. They learned that the most essential core value a location tracking app could provide was the value of 'Help.' They also learned that the value of privacy becomes secondary for seniors the moment they (the customers) get real core value back in the form of ‘Help’ (in exchange for sharing their location data). The core value of ‚Help‘ could be provided in this case through value qualities such as: quick access to sales associates, improved orientation in the store, or time savings. Value-based engineering therefore aims to let innovation teams always interact with stakeholders first (in this case, the seniors) to think about the context specific value dynamics (core values and value qualities) of an SOI and to do so free of value lists. This engagement with the directly experienced or imagined operational reality of an SOI allows to elicit, question, and recognize what might be of value to future direct and indirect stakeholders. And it is precisely these values to which an SOI should then be primarily aligned during development.

This approach of an unbiased examination of the given value diversity of a SOI context is in line with the recommendations of many thought leaders, such as Ikujiro Nonaka, who assigns great importance to the anticipation of values in order to successfully innovate [7]. It also aligns with the work of Batya Friedman, whose 'Value Sensitive Design' approach has accumulated two decades of experience in identifying contextually relevant values and incorporating them into engineering systems [8]. And it aligns with what Design thinkers would call “emphasizing”.

Nevertheless, two critical questions may be asked in response to this case: First, what does a value like ‚Help’ in this location-tracking example have to do with ethics? Second, does the value elicitation phase of the VbE or IEEE 7000™ de-prioritize such a legally and morally recognized value as privacy behind a practical value such as customer ‚Help‘? The following sections will answer these questions.

III. MATERIAL VALUE ETHICS AND MORAL PHILOSPHY ARE FOUNDING VALUE-BASED ENGINEERING

Max Scheler's material ethics of value, on which VbE is founded, considers everything surrounding us, other people, nature, technology, relationships, or activities as potential carriers of value [9]. Things like a voice assistant or a location tracking application can become *bearers* of value if they embed technical configurations that enable these. The configuration of the voice assistant dialog, for example, is the prerequisite for either the value of compassion to be revealed to a child user or the value of toughness. For the senior in the fashion store, the value of ‘Help’ may unfold when she looks at the location app on her smartphone and sees a big green button there that allows to summon staff. So the technical configuration is a *value disposition*. It is created by the system developer, and is available as a potential for value qualities and core values to unfold. The result for the ecosystem of stakeholders is positive. There is enrichment on both sides: those seeking help and those helping. Those interacting with speech assistants and those building it. But exactly this, *the creation and promotion of a positive ecosystem for involved direct and indirect stakeholders*, is the concern of VbE in its following of Scheler’s material value ethics [9]. Thinking

about 'value bearers', 'value dispositions', 'core values' and 'value qualities' allows to think about a system design from the beginning in a value-strategic and thereby ethical way. Ethics means to think, to perceive and to design the environment as value-laden.

A. Use of Moral Philosophies for the Exploration of Values

How do you ensure that a value strategy is really ethical? Especially when, for the reasons mentioned above, one does not initially resort to predefined, institutionally approved lists of values or established norms? In VbE or when following the IEEE 7000™ standard three recognized moral theories are expected to be applied in context. With reference to the respective SOI and with an initial Concept of Operation in mind, three questions are derived from utilitarianism [10], virtue ethics [11], and duty ethics [4]:

- **utilitarianism:** what human, social, economic, or other values are affected, positively or negatively, by the SOI?
- **virtue ethics:** what is the long-term impact of the technology on the character of the affected stakeholders?
- **duty ethics:** what personal maxims or value priorities does the project team see affected by the service that the project team members believe are so important that they want to preserve them in society?

These three questions are elaborated not only by the project team directly responsible for building the SOI, but by a broad group of stakeholder representatives who are involved in the development of the value strategy and whose role it is to take a critical and holistic view of it. The discourse between stakeholders as well as their selection should meet requirements such as those envisioned by Jürgen Habermas in his Discourse Ethics [12].

If a culture in which the SOI is to be used has a specific ethical orientation embedded in it that goes beyond these three ethical theories, then the IEEE 7000™standard encourages adding additional questions for that culture, which are grounded in its philosophy or spiritual tradition.

The result of such a guided reflection on values is that even for simple SOIs, a relatively broad spectrum of values is identified. Across three case studies, we empirically observed that each stakeholder identified an average of 16-19 values [13]. This adds up to a significant number of values for larger stakeholder groups even if some of the values found will overlap. For one of the three cases, a Viennese telemedicine start-up, innovation management students identified a total of 54 unique value violations that could arise from the platform, as well as 63 positive values promotable [13]. This large spectrum of values, both positive and negative, makes one aware of how ethically fine-grained and sensitive a technology can actually be when scrutinized with the help of a larger critical stakeholder group. Uncovering the value potentials, however, does not only sharpen ethical sensitivity. For investors, it also makes clear how risky a new technology investment can actually be. And for product management, the positive value potentials open up a host of indications for refining the "value proposition" at the center of their business model.

To cope with the complexity of such a value spectrum, VbE discerns so called "core values" and "value qualities". *Core values* are values that are repeatedly described or even directly named by the assembled stakeholders and which carry special weight. As core values they should be formulated positively and be *intrinsic* in nature. This means that they are desired for their own sake and that even across cultures there is seldom any doubt as to what they are supposed to be good for (for example: friendship, community, dignity, knowledge, security, etc.) [14]. Complementary are the *value qualities* instrumental to the core values. These also result in their initial form from the stakeholders' descriptions. For example, if a stakeholder is concerned about the privacy of his voice assistant, he might say that he does not want the voice assistant's knowledge about the user to be sold or that there should be no unauthorized recording, that data security must be guaranteed, etc. Such stakeholder statements demonstrate the contextual "qualities" of the core value (privacy), which should therefore be respected in the SOI's design. In the IEEE 7000™ standard value qualities are called "value demonstrator".

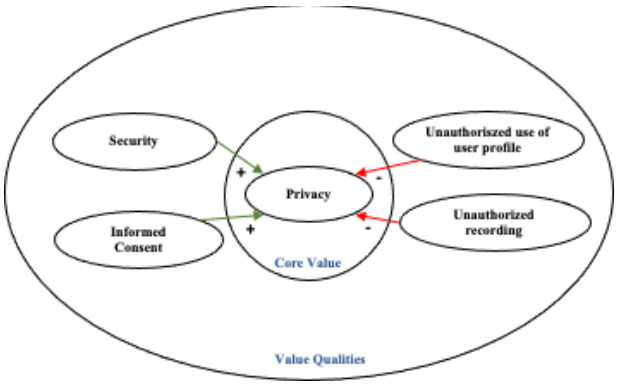

Fig. 1. Rough example of an initial core value cluster

Finally, it should be added that the value qualities collected in a stakeholder process are usually not complete. Stakeholders have intuitive access to values. In Scheler's words, they recognize them through their "value feelings". For example, stakeholders may indeed feel fear of losing their privacy when dealing with voice assistants [15]. But feelings are usually not enough to fully conceptualize a core value. Figure 1 clearly shows this. Anyone familiar with the privacy principles considered in the EU's General Data Protection Regulation knows that there are significantly more qualities than stakeholders come up with bottom-up. For this reason VbE engages is what Value Sensitive Design has coined "conceptual analysis" [8]. In a separate analysis, VbE foresees the completion of value qualities in line with the law or the philosophical literature.

B. Using Moral Philosophy to Prioritize Values

In the case of the telemedicine start-up, 13 original core value clusters could be identified with many value qualities attached to these. Which ones should be considered? VbE and IEEE 7000™ do not envision that core values will be pitted against each other; that is, that they will be treated as a trade-offs. Instead, the clusters are presented to executives and stakeholder representatives to rank them in terms of their importance to the SOI and the company. This ranking considers the following criteria: (p. 41 in [6]):

1. „Stakeholders agree that the SOI is good for society and avoids unnecessary harm.

2. The organization does not use people merely as a means to some end.
3. Organizational leaders can accept responsibility for the value priorities chosen according to their own personal maxims.
4. The organization respects its own stated ethical organizational principles if there are any.
5. The organization can commit to the value priorities in its business mission.
6. The environment is maximally preserved
7. The organization considers existing ethical guidelines."

For the philosophically trained, reading these seven criteria immediately reveals that the second and third criteria are borrowed from Kant's duty ethics [4]. The fourth criterium takes into account existing lists of value principles developed by an organization or industry. It has been criticized that such lists, which are often developed by CSR departments, do not find much entry into lived corporate processes. However, through VbE and IEEE 7000™ they are called upon and can influence which SOI values should be prioritized.

Finally, it should be ensured that the value priorities favored by stakeholders conform to external value expectations. External value expectations are, for example, values enshrined in law, industry commitments or international agreements on ethical behavior, such as the United Nations Convention on Human Rights. So if a stakeholder team has ranked the value of customer 'Help' (in the fashion house example) before that of customer privacy, then aligning with the legal and political externals at this point ensures that privacy is finally ranked before comfort or help.

At the end of value exploration and prioritization, the organization and all parties involved have a very accurate view of the good and bad potentials of the SOI. If it turns out that value violations outweigh or that positive values are ultimately unpromising compared to the damage that can be done, the project team and management should decide not to invest in the system. Unlike IEEE 7000™, VbE encourages the active consideration of foregoing a technology for ethical reasons. Progress does not spring from bringing technologies into the world that may cause more harm than good for humanity.

IV. FROM PRINCIPLES TO PRACTICE

If a company decides to invest in an SOI taking into account its value priorities, the next question is how the value strategy can be systematically incorporated into the system design. To this end, each "core value/value quality" tuple is translated into a so-called *Ethical Value Requirements* (EVR). EVRs are not technical system requirements, but more precisely specified aspects of the socio-technical SOI. EVRs need to be put in place in order to bring a value quality into fulfillment. Thus they are requirements one could say, which serve as a target line for the ethically founded system design. For example, 5 EVRs for value quality of 'informed consent' (shown in Fig 1) are: the *understandable* and *honest, complete, easily accessible* information about the data processing actually taking place, and the *easily manageable* user option to prohibit this data processing beyond what is strictly necessary. An SOI, such as a voice assistant, can only be considered ethical if it meets these EVRs.

EVRs are the starting point for deriving system requirements. However, the above examples show that EVRs are not only of technical but can also be of organizational nature. The comprehensibility, honesty, and completeness with which an informed consent is sought is not a technical issue. Instead, the company's management might be required to openly and plainly explain its participation in so-called personal data markets to its customer.

This contrasts with a technical EVR that calls for ensuring the security of a voice assistant. Here, the tuple "core value: privacy/value quality: security" calls for technical engineering so that communication on the front end and data processing on the back end are designed to be secure. And because a lack of security could cause significant harm to the business and the user, technical departments should have a vested interest in not overlooking security-critical requirements in system design. With this in mind, VbE advises adopting a rigorous risk-based design approach. Rigorous risk-based design, as understood by VbE is a method for identifying system requirements through threat and vulnerability analysis in a way already standardized for safety, security and privacy [16]. In VbE, the design methodology is applied to all those values that are of particularly high protection demand. The "core value/value quality" tuple is placed at the center of the protection analysis. For all threats to values, appropriate system controls are chosen in line with protection demand. Controls can be technical or organizational in nature. The technical system controls identified in this way are then included in the general pool of functional system requirements. As a result an ethically and functionally aligned product roadmap is created.

A. The link to agile forms of system development

How does VbE deal with value qualities whose EVRs are not of an organizational nature and which are also not so important in their nature that they would have to be subjected to rigorous risk-based design? In fact, many EVRs are of exactly this kind of hybrid nature. The examples presented above have shown this already: A voice assistant may convey the value of 'compassion'; a tracking app the core value of ‚help'. Such values must be embedded in the system through respective technical value dispositions, but the nature of these values is such that rigorous risk-based design does not seem necessary. In fact, designing a voice assistants value of compassion with the help of rigorous risk-driven design method (such as the one NIST recommends for the value of security [16]) seems literally like 'shooting methodical cannons at sparrows'. Against this background, VbE offers a third way to anchor EVRs in the technical system design. In fact, nothing other than well-known, classical iterative and agile system development processes are used. A first viable product - perhaps a still unspecific prototype – may be used and stakeholder representatives discuss it. The key here is that the iterative development of the prototype is ethically guided by EVRs. Thus, the project team does not seek "tabula rasa" for user needs, but aims for values elicited with the help of moral philosophy. And prior to a sprint, EVRs have been run through a simple threat-control analysis. Figure 2 roughly illustrates the flow from principles to practice as promoted by VbE. (Note that IEEE 7000™ does not distinguish between a rigorous or agile derivation of system requirements).

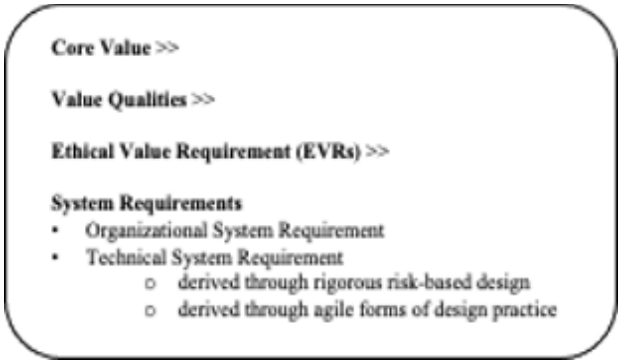

Fig. 2. Rough flow of constructs for ethical system requirements derivation

V. CONLUSION AND CHALLENGES FOR VALUE-BASED ENGINEERING

The description of VbE and some parts of IEEE7000™ makes clear that ethical engineering is not an isolated corporate process that can be outsourced to a department such as Compliance Management, Risk Management or the CSR Department. Ethics cannot be delegated or outsourced to any corporate niche function. For the 'good' to succeed, ethics must be woven into all relevant innovation processes as an end-to-end principle. VbE therefore envisages using value ethics and moral philosophy as tools in early phases of product development to influence corporate strategy and product design. This ultimately means influencing the 'value proposition' at the heart of the business plan. Executives, product managers and technical departments must work closely together to achieve this, which is not necessarily a given today.

Similarly, the company, which is responsible for the SOI from the customer's point of view, must involve all relevant service partners in the VbE to ensure the consistency of the value principles. The ecosystem around the SOI (the so-called "system-of-systems") must be on board to avoid risking unexpected and uncontrollable value undermining. This again means that SOI operators may have to give up some convenient supply chain partners. The product roadmap is then influenced just as much by value priorities as by purely functional system requirements. This in turn means an investment in an intangible "good" that not every capital provider likes to see today. In addition, the whole engagement takes time, not only to work out the ethical value strategy described here, but also to implement it in system terms. In short, VbE requires a fundamental rethinking of the way we approach systems development today. It requires more time, more money, more cooperation and care for stakeholders. That none of this is easy - whether standardized by IEEE or not - is abundantly clear to the author of this article. But it is also clear that a more humane, social, and good technology will not be possible without a radical rethinking of our innovation processes. Anyone who is committed to a 'better' world - and this is ultimately the basic concern of the IT ethics movement - must not believe that they can simply carry on as before, just ticking some ethical principle boxes.

The question is, of course, whether society, companies, administrations and politicians are ready for this change today. In the telemedicine case study described above, the CEO was not enthusiastic about facing all the negative value potentials of his platform, let alone spending money to avoid them. Accepting criticism constructively and investing money in ethics when you need to save cost is quite an ambition in the current economic climate. Therefore, for it to bear any fruit, VbE with IEEE 7000™ should ideally accompany the early stage of system development; at a time when the value strategy and thus the business model are not yet fixed. If the method is used in a brownfield situation, there must be a high degree of willingness by managers to rethink the company's business mission.

Finally, in addition to the maturity, willingness and ability to accept feedback, and the courage to seek it from outside critical stakeholders, there is another challenge: Ethical success cannot be measured precisely. This is not only the case because capturing what has not happened is not possible (like an avoided security breach), but also because what is 'good' is not precisely measurable. How should one measure love, friendship or dignity? Looking at a platform example like Wikipedia: How can one proof that it promotes the value of ‚knowledge'? Even when it comes to such an indispensable platform critical voices have discussed whether it always provides knowledge. Values ultimately have the nature of eternal ideas, similar to those portrayed by Plato in his allegory of the cave. In other words: if we continue to follow the current discourse, in which only the measurable is ascribed a reality, then value-based methods will have a hard time in corporate practice. Thus, VbE also heralds a cultural change for management. It is no longer about Peter Drucker's view that only what is measurable is manageable, but about the insight that only what has a shared value basis and trust is worth striving for, even if this is not measurable.

ACKNOWLEDGMENT

I want to thank the IEEE 7000 CRG Group Members for participating so actively in the development of the standard and thereby also informing the Value-based Engineering methodology. I want to acknowledge the support of my Ph.D. students Kathrin Bednar and Till Winkler who accompanied me in this research endeavor.

REFERENCES

[1] A. Jobin, M. Ienca and E. Vayena, "The global landscape for AI ethics guidelines," Nature - Machine Intelligence, 1, 2019, pp. 389–399.

[2] B. Mittelstadt, "Principles alone cannot guaranteee ethical AI," Nature machine intelligence, 1, November 2019, pp.501-507.

[3] P. Aronson, and J. Duportail, "The quantified heart" Aoen, 2018.

[4] I. Kant, "Kant's Werke" Suhrkamp, City, 1786.

[5] B. Berenbach, and M. Broy, "Professional and Ethical Dilemmas in Software Enginneering," IEEE Computer, 1, 2009, pp.74-80.

[6] IEEE 7000 - Model Process for Addressing Ethical Concerns During System Design. IEEE Computer Society, City, forthcoming in 2021.

[7] I. Nonaka, M. Kodama, A. Hirose and F. Kohlbacher, "Dynamic fractal organizations for promoting knowledge-based transformation – A new paradigm for organizational theory," European Management Journal, 32, 2014, pp.137-146.

[8] B. Friedman and D. G. Hendry, "Value sensitive design: Shaping technology with moral imagination" MIT Press, 2019.

[9] M. Scheler, "Formalism in Ethics and Non-formal Ethics of Values: A New Attempt Toward the Foundation of an Ethical Personalism," Northwestern University Press, USA, 1921, 1973.

[10] J. S. Mill, "Utilitarianism" Penguin Books, City, 1863/1987.

[11] Aristoteles „Nikomachische Ethik" Reclam Verlag, City, 1969.

[12] J. Habermas, "The Theory of Communicative Action Vol 1: Reason and the Rationalization of Society" Heinemann, London, 1984.

[13] K. Bednar, and S. Spiekermann, "On the power of ethics: How value-based thinking fosters creative and sustainable IT innovation" Working Paper: https://epub.wu.ac.at/7841/, 2020.

[14] T. Ronnow-Rassmussen, “Intrinsic and extrinsic value.” Oxford University Press, City, 2015.

[15] E. Görnemann, and S. Spiekermann, “Moments of Truth with Conversational Agents: An Exploratory Quest for the Relevant Experiences of Alexa Users.” City, 2020.

[16] NIST NIST 800-53: Security and Privacy Controls for Federal Information Systems and Organizations. U.S. Department of Commerce, City, 2013.