CAN ETHICS ENHANCE CREATIVE DESIGN ACTIVITY?

Sekiguchi, Kaira (1); Hori, Koichi (1,2)

1: University of Tokyo; 2: RIKEN Center for Advanced Intelligence Project (AIP)

ABSTRACT
This study exhibits that there exists generative ethics in which ethical thinking will allow us to change our perspectives to consider the artifacts’ social effects in design, thereby enhancing our activity to generate novel and practical design ideas.

As an example of practicing the generative ethics, a case for addressing the increasing requirement to properly introduce “artificial intelligence” (AI) systems in society was considered. We applied the ethical design theory to promote the practice of ethical AI design by engineers. To achieve this, we implemented a creativity support tool that can be used based on the knowledge base of AI ethics.

To confirm the validity of the theory and the tool, we conducted user experiments in which the AI research students had to consider the effects of their own research projects with using the tool. We could confirm that the tool actually induced the users to consider social impacts. Some students revealed in response to the questionnaire that the experiment provided them with an opportunity to reconsider their own research theme.

In this study, the ethical design theory and tool will be briefly reviewed, and the experiments will be discussed.

Keywords: Design theory, Ethical design, Creativity, Support tool, Knowledge management

Contact:
Sekiguchi, Kaira
University of Tokyo
Department of Aeronautics and Astronautics
Japan
kaira@dfrome.com
1 INTRODUCTION

The impact of science and technology on society has generally been investigated in terms of applied ethics, including in terms of environmental ethics, bioethics, ethics of technology, etc. For example, ethics of technology consider the impacts of the Internet in society. Meanwhile, because of the continued existence of social issues, the responsibilities of science and technology to society and the need to address the demands of society have been recognised as mentioned in Yoshikawa (2003). To positively promote this realisation in engineering, focusing on the negative aspects of science and technology and the restriction of activities is considered to be insufficient. Even if there is no opposition to the regulatory nature of ethics from the field of engineering, engineers’ active participation in ethical investigation cannot be expected. Rather, it is important to exhibit that the use of ethical practices by engineers is useful to improve the value of the output of their research and development. Here, these ethical practices align with the engineering way of thinking which demands a scientific explanation of the creation and realisation of new technology. Therefore, we call such ethics ‘generative ethics’.

Thus far, to deal with ethical issues from the engineering way of thinking, we have redefined design perspective and design description method Sekiguchi et al. (2010). We further reviewed the perspective and method and proposed them as the ethical design theory; in addition, we implemented and provided a creativity support tool based on such theoretical knowledge Sekiguchi (2017); Sekiguchi and Hori (2018a). This support tool was originally intended to support the appropriate introduction of ‘artificial intelligence’ (AI) ethics into engineering society, which has become a massive problem in an recent years. Further, we have exhibited the relevance of the ethical design theory and the tool by describing use cases. Additionally, we adopted different logic and a case to demonstrate the validity of the theory and of the improved version of the support tool Sekiguchi and Hori (2018b). One of the actual examples of the usage of the tool is a redesign of the drone delivery service, which was encountered with the help of an expert on public law and an expert on AI ethics/AI Sekiguchi et al. (2018).

As mentioned above, the proof of the validity of the ethical design theory and of the tool has already been obtained based on considerations by the authors and a third-party. Therefore, to prove the validity more objectively, it must be stated as a next step that the same effect can be obtained by practising only by a third-party. Thus, we have conducted user experiments using the aforementioned support tool. The goal of this paper is to describe the experimental evidence that supports the existence of generative ethics: ethics that enhance creative design activity. In summary, the contributions of this paper are: (1) to describe obtained evidences that show the existence of the enhancement in which ethics support creative design activity; (2) to describe obtained evidences that show how subjects introduce ethical perspective that is useful for the enhancement; and (3) to provide a way of using our support tool for a reproductive experiment.

In the subsequent sections, after reviewing the ethical design theory and the outline of the creativity support tool as well as the positioning of our work among related works, we describe and discuss the settings and results of the experiments, and describe the conclusion and future works.

2 ETHICAL DESIGN THEORY

We first review the ethical design theory, which comprises of perspectives and a description method.

2.1 Design from the ethics level

First, to deal with ethics according to engineering practices, we have to redefine the design perspective. Traditionally, in engineering, engineers consider the artifacts in a hierarchical manner Simon (1996); Yoshikawa (1979, 1981). Before beginning with our research study, the hierarchical representation of artifacts was discussed up to the interaction level Nakakoji (2007), as depicted in (H) in Figure 1. Further, we positioned the ethics level as one that corresponds to the whole society at the top of this hierarchy, as depicted in (1) in Figure 1. Here, ‘ethics’ means a system of social values and ‘social values’ are indicated using social scale values such as ‘human dignity’, ‘rights’, ‘freedoms’ and ‘cultural diversity’. For example, these values are introduced as human values in Asilomar AI Principles (Future of Life Institute, 2017). The higher the hierarchy, the more its range of influence will be. Therefore, it is reasonable to position ethics at the top level of the hierarchy. Further, if the phenomena that are observed at this level correspond to that of the ethics of technology, then it is possible to consider the orthogonal axis that corresponds to engineering ethics, as depicted in
(O) and (2) in Figure 1. In the orthogonal axis (O), the reasons in (P), actions in (H), and results in (E) that are obtained by the designer can be denoted sequentially as (T).

We can also describe these three-dimensional perspectives in a simpler manner by using the hierarchy (H) and the projection images (P’) and (E’), as depicted in Figure 2; thus, we have adopted these perspectives in our creativity support tool.

Figure 2. A simplified version of the hierarchical and orthogonal representation of design ideas.

2.2 Design with discourse

To perform designing by considering the ethics level, we introduced a method for describing design ideas that we called ‘design with discourse’. One of the most important rules of design with discourse is to describe the changes that will be caused by the designed artifacts.

This rule corresponds exactly with the understanding that the designing activity generates desirable changes that cannot occur in the natural state. For example, the famous definition of design by Herbert A. Simon which says ‘Everyone designs who devise courses of action aimed at changing existing situations into preferred ones’ Simon (1996).

In particular, the description rule combines the three descriptions as follows and as depicted in Figure 3: (1) since A is a personal reason, I/we generate a design that will change B to C in the hierarchical representation of artifacts; (2) if B is changed to C at the parameter level, then D will change to E at the target level in the hierarchical representation of artifacts; and (3) if D is changed to E in the hierarchical representation of artifacts, then F will change to G as effects on me/us. Hence, these descriptions will form a tree of design ideas when they are connected one by one:
3 OVERVIEW OF DFROME

In this section, we review the creativity support tool called Dfrome, which is an abbreviation for ‘design from the ethics level’.

3.1 Aims

Dfrome is considered to be an organic and dynamic creativity support tool. Here, ‘organic’ indicates that different relations among different AI ethics and technologies are properly signified as different; ‘dynamic’, on the other hand, indicates that the tool dynamically adopts new issues and helps engineers to think in contexts that are relevant to their project. The original objective of Dfrome was to fill the gap that exists between AI engineers and the discussion of AI ethics, and to support the AI engineers in utilising the results of AI ethics in daily research and development.

One of the most important functions of Dfrome is the recommendation of scenario paths. This function involves the presentation of the scenarios that should be considered by the user in accordance with the query submitted by the user. There are two options that users can choose for their computational processing. The first option is to choose whether the results will be the scenarios of impact on the outer environment (which is upward in the hierarchy) or the scenarios of related element technologies (which is downward in the hierarchy). The other option is to choose whether the results will be the scenarios that are considered to be relevant in the context of the query or the scenarios that diverts the user’s thought from the context of the query.

3.2 Components

The latest version of Dfrome comprises the following four components Sekiguchi and Hori (2018b):

1. an editor for editing the description of design ideas in accordance with the ethical design theory;
2. a cloud environment for saving and sharing the edited ideas as contents of Dfrome’s knowledge base;
3. an investigation engine for purposes, including the recommendation of scenario paths; and
4. a browser for casual users to browse various scenario paths.

In this experiment, a browser is used as the interface instead of an editor because we assume the presence of casual users in this experiment. An overview of the browser is shown in Figure 4. Based on design with discourse, when a user agrees to the terms, enters a query in the text field, selects two options and presses an execution button, a scenario is presented as a connection of description.

An overview of the processing of scenario path recommendation of Dfrome’s browser is shown in Figure 5. The user is supposed to query a sentence by setting the aforementioned two options. Then, Dfrome will calculate to order the result T paths that are selected from candidate M paths which are divided from the stored N trees and will visualise the results one by one.

3.3 Knowledge base

In the knowledge base of Dfrome, sixty nine trees are accumulated when this experiment was performed. For example, it includes trees of recent principles and guidelines of AI ethics such as Asilomar AI Principles (Future of Life Institute, 2017), trees of scenario analysis (The Conference toward AI Network Society, 2017b) which include risk scenarios and drone studies as described in Sekiguchi and Hori...
4 RELATED WORKS

An example of similar research is Ibo van de Poel’s Ethics and Technology 2.0. The objective of his research is to update the ethics of technology and he described these three points: (1) from reactive to proactive (not after a technology has been developed but already in the early phases); (2) constructive approach (ethics not as a break but to the development of technology); and (3) not only prevent harm but do good (better technologies that contribute to a better society) van der Poel (2015). Thus, he introduced the value hierarchy as one of the ‘value sensitive design’ (VSD) tools and methods, whose hierarchy is composed of (i) values, (ii) norms and (iii) design requirements.

The broadest difference of his work from our hierarchy is that his hierarchy is not based on the hierarchical representation of artifacts. Therefore, in value hierarchy, it is unclear as to which level in the hierarchical representation of artifacts does a particular value correspond to. Briefly, it is not guaranteed that the social value is described in its values. For example, the fact that a certain material is light can be considered to be one of the values; however, this value can be described at the material level within the hierarchical representation of artifacts. On the other hand, our hierarchy can allow us to clearly recognise whether the value corresponding to the ethical level is described because it absolutely defines the ethical level at its top. Additionally, our hierarchy can allow us to clearly recognise the level at which a
gap of ideas exists and seamlessly fill the gap. Therefore, it can be stated that our hierarchy is suitable for helping engineers in developing the ethical design in a clearer and more practical way as well as based on a perspective familiar to engineers (i.e. hierarchical representation of artifacts).

Furthermore, although van de Poel’s work distinguishes between the ethics of technology and engineering ethics, it seems that Ethics and Technology 2.0 mainly covers phenomena corresponding to the ethics of technology. However, our three-dimensional representation that has been organically defined combines both ethics of technology and engineering ethics. (Additionally, a comparison with value dams and value flows of VSD has also been described in a previous study Sekiguchi and Hori (2018)).

In summary, based on our three-dimensional representation, the knowledge base that is obtained will be able to accumulate high-quality data. To the best of our knowledge, this study is the first that presents experimental evidence of the validity of the ethical design theory and Dfrome for enhancing creative design activity. For example, it was experimentally shown for the first time that it is probable to change engineers’ perspective to consider a society with Dfrome, that is, it is difficult if the hierarchical representation of artifacts is not adopted as in the case of ordinary web search.

Then, Finke et al. suggested that proper constraint plays a vital role in creative activities Finke et al. (1992). Although the studies of Finke et al. were focused on elements in creative activities, our experiments demonstrated that the suggestions are also valid in a practical activity i.e. the ethical design.

5 EXPERIMENTAL SETTINGS

One of the goals of this experiment was to verify whether the practice of generative ethics is possible even if Dfrome is used only by a third-party AI engineer. Specifically, we observed that using Dfrome leads to new awareness for these users and may result in them to update their own research themes.

The procedure for this experiment is described as follows: (1) explain the task and how to use Dfrome (about 5 minutes); (2) do the task by the user himself/herself (Total 40 minutes = 20 minutes × 2 frames); and (3) conduct a questionnaire and a hearing. In subsequent subsections, details of environment, subjects, task, questionnaire and hearing are described.

5.1 Environment

The experimental environment and its actual image are as depicted in Figure 6. Only one participant took part in one experiment. Dfrome runs on a laptop that records the operation screen during the experiment. The state of the subject was also recorded to the extent possible.

The experiments were basically conducted in Japanese. Dfrome is only available in English; any language can be used for web search and task paper sheet.

5.2 Subjects

Four master course students at the AI Lab at the University of Tokyo participated in the experiment. A brief overview of each theme category is presented in Table 1.
Table 1. Brief overview of each user’s research theme

<table>
<thead>
<tr>
<th>Experiment order</th>
<th>Id (in this paper)</th>
<th>Grade</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Master in 1st grade</td>
<td>Localization, Drone</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>Master in 1st grade</td>
<td>Data analysis</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>Master in 2nd degree</td>
<td>Deep learning</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>Master in 1st degree</td>
<td>Drone</td>
</tr>
</tbody>
</table>

5.3 Task

An A4 paper sheet was provided of which the following text is written at the top (Left side of Figure 7):

What kind of impact will the result of your research and development have on the stakeholders when it is realised? Please fill in the blank freely.
Please speak out loud as you think. Internet environment is available. The time is 20 minutes for 2 frames. Please use the tool only for one frame (Explanation will be held).

Figure 7. An Image of a task paper sheet and a questionnaire paper sheet (in Japanese)
As shown here, there are two frames for each experiment: one with Dfrome and the other without Dfrome. Before explaining the frame using Dfrome, we briefly explain Dfrome. Although the explanatory was actually separated into two pages, it was essentially the same as that shown in Figure 4. Additionally, the following two hints were also provided.

Let’s query an abstract of your research paper.
If there are insightful descriptions in the recommended scenario, let’s query them next.

In the experiment, observe that the use conditions of Dfrome were assigned according to the order of participation. Specifically, subject A and C used Dfrome in the first frame; B and D used it in the second. Since the order of the experiment was determined by external factors such as the schedule of the subject, the use conditions of Dfrome can be regarded as almost randomly allocated.

5.4 Questionnaire and hearing
The questionnaire asked whether there was any novel awareness, whether this experiment provided an opportunity for the participant to reconsider his/her own research theme, etc. (Right side of Figure 7). After that, a hearing was conducted based on the task paper sheet, the answers to the questionnaire, the record of the operation screen, etc.

6 RESULTS OF EXPERIMENTS AND DISCUSSION
In this section, the experimental results are presented and discussed. It is worth noting that quoted sentences stated below were translated from Japanese into English by authors and parts with parentheses below are comments by authors. Further theoretical discussion is also described in the next section.

6.1 General results
All the subjects used mainly Dfrome, when it was available. Three of them supplemented it with Google mainly only for confirming the meaning of some words or searching sentences used as a sample query. Therefore, Dfrome is confirmed to have functions basically necessary for doing the task. In the absence of Dfrome, all the subjects mainly used Google for searches.

6.2 Changing the perspective in design
The two subjects that first used Dfrome naturally accepted ethical thinking. For example, subject A in Table 1 answered in the questionnaire ‘Although I had thought about benefits up to now, I have not thought about risks, so I realised awareness about the risk’. This is an example of accepting ethical thinking of broader issues about the influence of technology, including the risk. Further, subject C commented at the end of the frame using Dfrome that ‘(Time) was unexpectedly short and it was difficult because I had not thought much (this thing)’. This is also an example that he accepted ethical thinking different from the usual one.

On the other hand, the other two subjects, who started without Dfrome in the first frame seemed to obtain information based on the viewpoints that they originally had. For example, subject B commented in the hearing after the experiment as ‘Since I have thought a problem about uncertainty, I investigated this (even in Dfrome)’. Subject D answered as ‘Quite yes’ to the question which was ‘Was the way of thinking different between the first frame and the second frame?’ He said ‘The scope of thinking became wider’ and ‘In the first frame, I just described what I had thought so far and I sometimes searched and thought that “Ah, there was such research as well”’. However, in the second frame, it was first questionable whether there was such an unexpected scenario, but when I tried (using Dfrome), there were often such scenarios that I became aware of their feasibility’. In addition, in his task paper sheet, he also tried to connect his original idea of a drone to the novel viewpoint obtained by the task execution such as the idea of drone highway. Further, subject D answered to the questionnaire as follows ‘I have so far thought about technical aspects, however, I realised that thinking functionality will allow me to conduct my research in more self-disciplined manner and would like to consider functionality after this’. In this way, Dfrome succeeded in inducing the subjects to develop an ethical viewpoint as we aimed at the beginning. On the other hand, as Google mainly provides an information retrieval function, it turns out that it does not tend to support a change in the viewpoint i.e. the concept space of the user.
6.3 Reconsidering his/her own theme

Subject A and D answered in the questionnaire that this experiment likely provided an opportunity to reconsider their own research themes. For example, subject A answered ‘I have been paying attention to outdoor use (of Drone), but since I felt it was likely to be used indoors, I would like to think about indoor use and problems’. It was confirmed from the hearing and the operation screen record that the scenario that acted as a trigger was concerned with the customers’ privacy risk by the RFID data collection in the shopping mall. It was an unexpected use case where the user generates a new application by replacing the subjective artifact with his research theme’s artifact when receiving a risk scenario. He further said that he would like to think about its problem. Therefore, this use case provides an interesting example that ethical thinking leads to creative activity i.e. it strengthens the novelty of the user’s own research theme by allowing him/her to notice a novel problem setting. It is important that this consideration is practical because the way of this ethical thinking are based on the engineering perspective and grammar. Aforementioned comments on ‘functionality and subjectivity’ of subject D is also related to updating the research theme. By noticing the role of upper part of the hierarchical representation, the subject seemed to have realised that he was able to freely have the viewpoint (in the orthogonal axis).

The point is that these two subjects who were led to reconsider the theme were researching drones. With regard to drones, six trees were registered in relation to the collaboration Sekiguchi et al. (2018). This means that relevant contents were abundantly registered in Dfrome’s knowledge base. Therefore, we were able to confirm that it will be probable to support users to update their themes even with the current version of Dfrome if there are rich contents.

6.4 Other points confirmed

Thanks to the effectiveness of the description grammar, the efficiency in consideration was able to be improved. One typical example of this is a comment we received by subject A in his hearing saying ‘The tool was good because the scenario appeared in a form that can be immediately understood’. In fact, he seemed to struggle with Google because the record of the operation screen actually showed that he often scrolled the documents obtained via the Google search (in order to find where relevant descriptions exist).

Then, subject D mentioned in his hearing that ‘It was fun and I seemed to be talking with the supervising teacher (a professor of engineering) (when I see the scenario presented by Dfrome)’. This suggests that Dfrome was able to be equipped with a professional engineering way of thinking. Therefore, from this comment, we confirmed the probability that Dfrome is effective for teaching engineering students.

7 DISCUSSION

In general, we were able to observe how to create a new scenario by combining the new context each subject had and the comprehensive scenarios Dfrome accumulated. The point is to properly recommend a scenario by calculating the context of the user, that should be considered by the user. For example, subject A considered that a scenario on ‘privacy’ relates to his own research; subject D considered that a scenario on ‘safety’ relates to his own research. Furthermore, both of them tried to incorporate these perspectives into their own research and updated their ideas. This co-generation is the effect of the semi-automatic approach, and to the best of our knowledge, this is the first time that ethics were observed to interactively enhance creative design activity based on the ethical design theory.

In Dfrome’s database, scenarios that seamlessly connect from the technical levels to the ethical level are accumulated. When subjects read these scenarios, they construct the hierarchy in their heads and consistently observe their higher level objectives. These top levels correspond to the prerequisites that enable design at lower levels, that can be understood as the ‘hinges’ Wittgenstein et al. (1972) as discussed in Sekiguchi and Hori (2018a). Therefore, introducing ethical issues enables subjects to review their own research themes from the higher objectives, which encourages the exploration of wider design solutions. For example, through referring to the issues of ‘privacy’ by the researcher of RFID, subject A considered broader areas where privacy is problematic and found that some of these areas can also relate to his own research theme. Then, subject A thought that indoor (specifically, in a mall) usage also will be the scope of his own application.
8 CONCLUSION AND FUTURE WORKS

From the experimental results, we conclude the following: (1) the introduction of ethical viewpoints enabled users to generate more creative design by updating their own research themes and making them novel; (2) Dfrome was able to enhance the introduction although Google was confirmed to have a tendency to let the subjects deepen their own issues they already had; and (3) the experimental settings here were sufficient to observe such phenomena. Therefore, it can be said that these results support the existence of generative ethics.

As future works, broader contents must be accumulated in the knowledge base of Dfrome to be sufficient to lead all the users to review their theme. Therefore, we will continue to increase the knowledge base while implementing various collaborations. Furthermore, we will confirm the effects of Dfrome’s new functions such as those that automatically generate a draft of a tree from input sentences. Then, cause only subject D set a sentence of a recommended scenario as a next query although we gave a hint as depicted in Subsection 5.3, we will confirm the effects of a one-touch button to easily do this since its effects have been theoretically confirmed in Sekiguchi and Hori (2018b). Finally, we plan to conduct larger scale experiments and continue to provide Dfrome to allow others to conduct similar experiments.

REFERENCES


Future of Life Institute (2017), Asilomar AI Principles, Available at: https://futureoflife.org/ai-principles/.


Sekiguchi, K. (2017), Dfrome; Website for Design from the Ethics Level, Available at: https://www.dfrome.com (October 17–25, 2018).


ACKNOWLEDGMENT

The authors wish to acknowledge the four students who participated in the experiments. This work was partially supported by JSPS KAKENHI Grant Number JP18K18434.