AI & SOCIETY (2020) 35:51–71
https://doi.org/10.1007/s00146-018-0867-z

ORIGINAL ARTICLE

Organic and dynamic tool for use with knowledge base of AI ethics for promoting engineers' practice of ethical AI design

Kaira Sekiguchi[1] · **Koichi Hori**[1,2]

Received: 3 July 2018 / Accepted: 8 October 2018 / Published online: 16 October 2018

Abstract
In recent years, ethical questions related to the development of artificial intelligence (AI) are being increasingly discussed. However, there has not been enough corresponding increase in the research and development associated with AI technology that incorporates with ethical discussion. We therefore implemented an organic and dynamic tool for use with knowledge base of AI ethics for engineers to promote engineers' practice of ethical AI design to realize further social values. Here, "organic" means that the tool deals with complex relationships among different AI ethics. "Dynamic" means that the tool dynamically adopts new issues and helps engineers think in their own contexts. Data in the knowledge base of the tool is standardized based on the ethical design theory that consists of an extension of the hierarchical representation of artifacts to understand ethical considerations from the perspective of engineering, and a description method to express the design ideas. In addition, we apply the dynamic knowledge management model called knowledge liquidization and crystallization. To discuss the effects, we introduce three cases: a case for the clarification of differences in the structures among AI ethics and design ideas, a case for the presentation of semantic distance among them, and a case for the recommendation of the scenario paths that allow engineers to seamlessly use AI ethics in their own contexts. We discuss the effectiveness of the tool. We also show the probability that engineers can reconstruct AI ethics as a more practical one with professional ethicists.

Keywords AI ethics · Design theory · Creativity support · Hierarchical representation of artifacts · Knowledge base

1 Introduction

The importance of AI ethics is being increasingly recognized in recent years and many principles and cases have been provided by academic societies, foundations, administrative organizations (IEEE 2016, 2017; AI Network 2017a, b, c; FLT 2017), and so on.

However, AI studies have not sufficiently incorporated these results into their research and development; therefore, there remains a gap between ethical discourses and engineering practice. This is because AI ethics is too broad, and the amount of information is too large. Also, AI engineers have little time to understand and construct the relationship between AI ethics and their own research and development when they are involved in their investigations. Even if the AI engineers can find time, they often find it difficult to understand the relationship because the discourse of AI ethics is highly abstract.

In this research, we aim to fill this gap to realize further ethical AI and social values by supporting AI engineers to grasp ethical issues by extending their own research and development, and practicing an ethical AI design. Ethics can be classified as something related to social values or to personal beliefs. In this paper, we basically use ethics to mean that of social values such as freedom, quality, and human rights.

To realize these objectives, we implement and provide a knowledge base of AI ethics (which we call the "AI ethics library") for our organic and dynamic tool called "dfrome" (the abbreviation of "website for Design FROM the Ethics level") (Sekiguchi 2017). Here, "organic" means that dfrome deals with complex relationships among different AI

✉ Kaira Sekiguchi
kaira@dfrome.com

Koichi Hori
hori@computer.org

[1] Department of Aeronautics and Astronautics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[2] RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan

ethics. "Dynamic" means that dfrome dynamically adopts new issues and helps engineers think in their own contexts.

Dfrome clarifies the differences of structures among AI ethics and design ideas. It also presents semantic distances among them, and recommends scenario paths that seamlessly connect AI ethics for the extension of AI technology. In addition, it is expected that engineers who practice ethical AI design can provide feedback for AI ethics and make it more practical.

2 Related works

To support ethical AI design with a tool for use with knowledge base of AI ethics, we must first understand the characteristics of design representation. From the beginning of design study as an academic discipline, a hierarchical representation has been claimed to be essential for design activities (Simon 1996; Yoshikawa 1979, 1981) because the relationship between objectives and means in design correspond to the relationship between the higher and lower levels such as the system and the subsystem. Therefore, certain studies have used these hierarchical expressions as the knowledge base to support the design. For example, Chapman and Pinfold (1999) used the product model tree, which is a kind of hierarchical representation. Myung and Han (2001) used a hierarchy, which involved relationships among engineering system, design units, and subunits. Lee et al. used a hierarchy saying, "[I]t is necessary to classify five levels of knowledge in order to use relevant knowledge in a systematic way" (Lee and Han 2010).

The point here is that the hierarchical representation has helped design support systems and relevant design theories to deal with function decomposition by descending to lower levels; this representation has helped deal with the kind of subsystem to be used to realize some kind of function as a system (Chapman and Pinfold 1999; Myung and Han 2001; Lee and Han 2010; Afacan and Demirkan 2011; Tomiyama 2016). For example, this representation has aimed not only to present registered design solutions but also to support divergent thoughts by providing alternative solutions.

To ensure ethics in design, it is essential to investigate the essential objectives by returning to higher levels in the hierarchy. However, only some research on creativity support systems have dealt with higher objectives. According to Wang et al., only seven studies dealt with such "problem finding" approaches (Wang and Nickerson 2017). Currently, no tool for use with knowledge base of AI ethics is available that supports ethical AI design; therefore, our study covers these aspects.

In detail, such studies are absent because the propositions at higher levels are prerequisites for lower-level designs, and the necessity to doubt the design at higher levels does not always exist. For example, Afacan and Demirkan (2011) developed a support system to aid universal design. Here, we can expect that they dealt with the user's values in the universal design to realize such high objectives. However, the objectives that they presented related to system's requirements at the lower objectives, for example, "What space dimensions are" and "How the space is furnished". There are not enough studies on why these individual designs are necessary in the first place. For example, these designs were not considerations from the level of social values, for example, to support people's freedom-based activities or to ensure fairness. In this way, they did not investigate ethics even in the context of universal design.

Meanwhile, disciplines in ethics also deal with hierarchical expressions, but it decomposes tasks rather than functions (Spiekermann 2016). They differ in nature from the above-mentioned hierarchy familiar to engineers. The details will be discussed in Sect. 3.4.

Finally, we consider the new activities that account for ethical issues such as positively constraining factors (Finke et al. 1992) for performing a more creative design. We do not consider them as regulators or brakes for creative activities. This is because by returning to the higher objectives, we can find unexpected consequences of designed artifacts and the spread of design solutions, which are not easy to imagine.

3 Ethical design theory

To deal with ethics in the hierarchical representation of artifacts, we rebuild the hierarchical perspectives as "design from the ethics level" and a description method as "design with discourse". These perspectives and method are expected to enable the standardization of ethical design; they help the user and the tool to easily understand the syntax and semantics. Therefore, we call them the "ethical design theory".

3.1 Design from the ethics level

To investigate AI ethics, we first raise the hierarchical representation of artifacts. Then, we extend it to deal with the other dimensions of ethics that correspond to personal concerns such as personal value judgment.

3.1.1 Definition and position of the ethics level

Simon (1996) and Yoshikawa (1979, 1981) considered the hierarchical representation basically up to the system level, which corresponded to the user interface. For example, Simon emphasized the importance of the concept of the interface as: "The artificial world is centered precisely

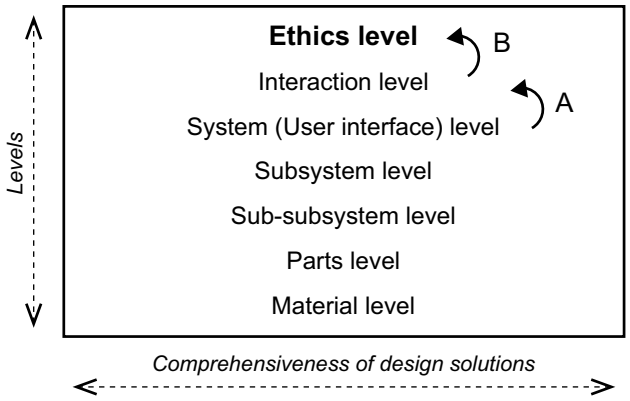

Fig. 1 Hierarchical representation of artifacts with ethics level (Sekiguchi et al. 2009)

on this interface between the inner and outer environment" (Simon 1996, p. 113).

Subsequently, the interactions became important. Therefore, Nakakoji stated: "While the term 'interface' makes people consider the character of an artifact as a *surface*, 'interaction' makes people consider the *time*, *flow* and the *change* that the artifact creates" (trans. by author) (Nakakoji 2007). Therefore, Nakakoji added the level of interaction as shown in (A) in Fig. 1.

Finally, to consider social relationships, we define the ethics level at which social values are expressed (Sekiguchi et al. 2009) and position it at the top of the hierarchical representation of artifacts, as shown in (B) in Fig. 1. Here, "ethics" is a system of social values and "social values" are indicative of social-scale values, e.g., "human dignity", "rights", "freedoms" and "cultural diversity" which have been introduced as human values in Asilomar AI principles. This kind of ethics corresponds to evaluating impacts of artifacts on the whole subject such as society, world, natural environment, etc., practiced in ethics of technology or environmental ethics. There is also another kind of ethics known as ethics of technician which will be dealt with in Sect. 3.1.2.

Although AI ethics that we rely on are biased towards values related to human natures including human rights (IEEE 2016, 2017; AI Network 2017a, b; FLT 2017), our theory and tool are not limited to this and we can include the values of other ethics, e.g., the knowledge base of our tool can support the values of "biodiversity" and "sustainability of natural environment" in environmental ethics. Therefore, our tool will broadly redefine AI ethics as the knowledge base of our tool increases.

As Fig. 1 shows, the hierarchical representation of artifacts has evolved to be able to handle the influence over the long term and a wider area. By expanding this trend, we have positioned the ethics level corresponding to social values whose change tends to be the final objectives in the hierarchy, for example, to make the world a better place. Technically, for a system in the hierarchical representation, a lower level is a component of a higher level. This means that a higher level must be wider than the lower level. Therefore, the ethics level that was defined to correspond to the broadest subject should be at the top of the hierarchical representation.

There is a probability that a hierarchical relationship may occur in each level and the pluralism of interpretation also seems to exist, especially, for higher levels. Therefore, the higher the level, the more important the way of humanities and lower the level, the more important the way of natural sciences is. The point is that designers can also ask the logic of the scenarios to the ethics level through the hierarchy; this scenario connected to the ethics level provides detailed definitions of the values of ethics each designer used in each context. This allows for more practical discussions of the abstract ethical values.

Here, we introduce an example of privacy that must be a typical value at the ethics level. If we prioritize privacy protection the most, it can be positioned at the top of the hierarchy. Contrastingly, we can also consider privacy as a means to protect individual freedom.

Figure 2 is an example of an e-commerce system considering privacy. Here, for the sake of simplicity, hierarchical values to be realized at each level are described by nodes and their purpose–means relationship is shown by arrows. When realizing anonymized processing of data as a subsystem (a), the system carries out e-commerce considering privacy at the system level (b). As a result, users can experience an e-commerce experience without concern for privacy issues at the interaction level (c). At the ethical level, this design does not conflict with the social value of privacy protection as (d) that can be an ultimate goal though others use it to protect the value of society's freedom as shown in (d′) and (e′) of Fig. 2.

The point is that these differences are useful to give new awareness among designers. In this paper, we will also describe a method to design from the ethics level, along with a tool to support its practice.

3.1.2 Dimension of personal concerns

The hierarchical representation corresponds to the general phenomena, such as natural science subjects, on which engineering is based. Therefore, we can add the other dimension of ethics, that is, for personal concerns.[1] This additional

[1] The details are described in our essay. Please refer to K. Sekiguchi, The fifth rule of "design with discourse" for the orthogonal representation of moral concerns in design from the ethics level, October 30, 2010. at http://www.ethics-level.com.

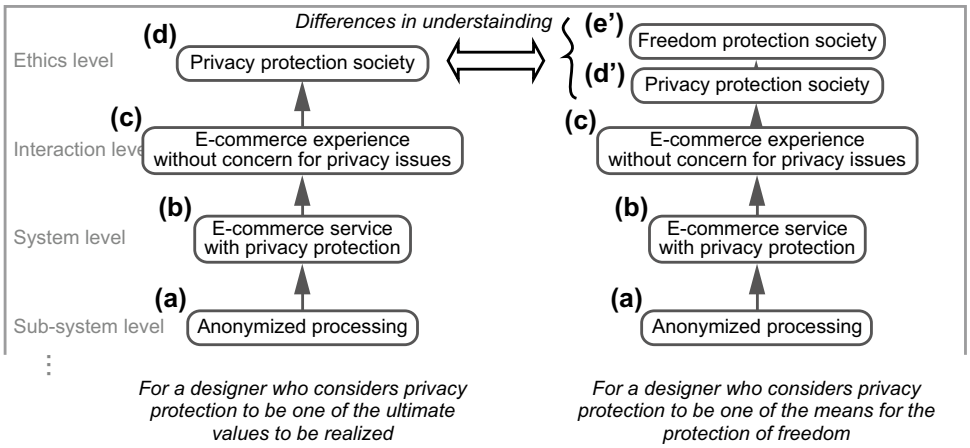

Fig. 2 Example of hierarchical values around an e-commerce system considering privacy protection in which differences of understanding occurs at the ethics level: **(a)** anonymized processing; **(b)** E-commerce service with privacy protection; **(c)** E-commerce experience without concern for privacy issues; **(d)** Privacy protection society; (d′) = (d); (e′) Freedom protection society

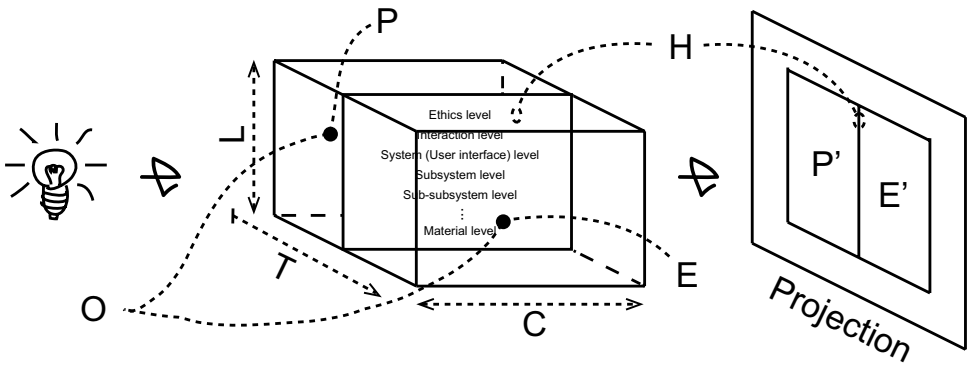

Fig. 3 Hierarchical and orthogonal representation of artifacts. *L* levels, *T* transitions of personal concerns, *O* orthogonal representation of personal concerns, *H* hierarchical representation of artifacts, *P* (*P*′) personal reasons, *E* (*E*′) effects on me

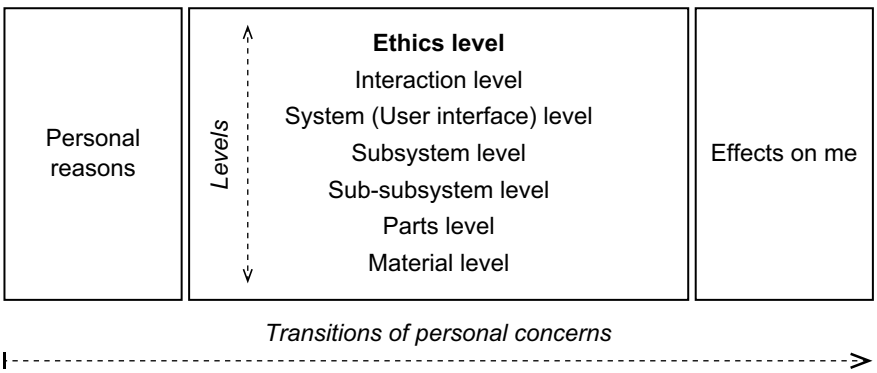

Fig. 4 Simplified version of the hierarchical and orthogonal representation of artifacts

dimension is orthogonal to the hierarchy of artifacts as O and T shown in Fig. 3 because we can consider each level directly, although changes must be generated from the parameter level that corresponds to the levels at hand in the hierarchical representation of artifacts.

As shown in Fig. 3, the two spaces can be placed on either side of the hierarchical representation to express personal perspectives as O. One space P is for personal reasons, and the other space E is for effects on me, as shown in Fig. 3. This is because the order and the direction of the arrow of T in Fig. 3 corresponds to the flow of time in the orthogonal representation of artifacts: a reason, an action, and an effect, in that order.

Then, we can simplify the three-dimensional representation using and positioning the projection field at the center, as shown in Fig. 4. We have applied this simplified version in this research.

In other words, the propagation of influence in hierarchical representation corresponds to objective causality which tends to be studied by the methods of natural science, whereas the orthogonal axis corresponds to subjective issues such as choice. For example, in the hierarchical

representation, the physical properties of biodegradability can be described at the material level. This material realizes parts return to nature after a certain period of time even without special disposal treatment. As the influence of this property spatially propagates, the value of this material will gradually change to a more socially recognized value such as disposable functions at the system level or sustainability of the natural environment at the ethics level.

Contrastingly, it is possible to make ethical choices for all levels, not just the ethics level. For example, a designer can choose biodegradable materials, because he or she considers that they are environmentally friendly. The foundation of this choice is that the designer believes that this material will achieve sustainability of the natural environment at the ethics level. The interpretation plurality discussed in the former subsection is based on how each designer understands the horizontally objective world from the perspective of this orthogonal axis.

At the open discussion during the 31st Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2017), one of the members of the ethics committee, Arisa Ema, introduced three major areas of ethics in AI: research ethics, AI ethics, and ethical AI (The Ethics Committee, The Japanese Society for Artificial Intelligence 2017). From our perspective, AI ethics is discussed at the ethics level; ethical AI is positioned as one of the artifacts designed in the hierarchical representation; research ethics indicates personal dimension issues that are orthogonal to the hierarchy. To the best of our knowledge, our research is the first to systematically deal with all three of these ethical dimensions as a whole.

3.2 Design with discourse

To design from the ethics level, we introduced a method called design with discourse, in which the lexicon and grammar were redefined to make the engineering general design theory to connect with the ethics. For the lexicon, we proposed using proper terms corresponding to each level (Sekiguchi et al. 2009). For example, terms used in the humanities including ethics discussed in Sect. 3.1.1, such as freedom, equality, human rights, and environmental sustainability, are significant at the ethics level. Here, we can refer to the importance of such conceptual investigation of values for value-sensitive design (Freedman et al. 2006; Miller et al. 2007).

Then, one of the most important functions of design grammar is to describe the changes between levels because design causes changes to realize values at higher levels (Sekiguchi et al. 2009). Therefore, we set one of our grammar rules as follows (also shown in Fig. 5):

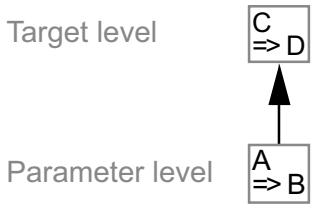

Fig. 5 Hierarchical grammar of design with discourse

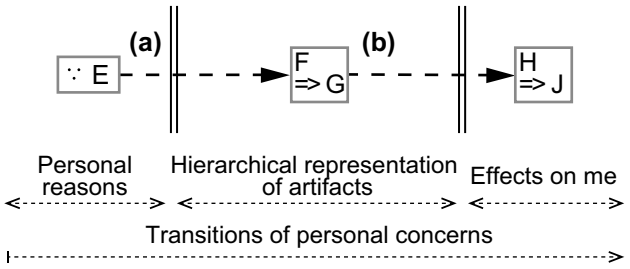

Fig. 6 Orthogonal grammar of design with discourse

- If A is changed to B at the parameter level, then C will change to D at the target level.

For example, we can describe the following: if the personal information is changed to be anonymized in the human–AI interaction, then the AI systems will change to firmly protect the users' privacy. By connecting these items of changes, the artifacts are designed to cause changes and realize values from the parameter level to the target level, for example, the ethics level.

Since we did not introduce this rule in Fig. 2, we described each node as an entity rather than a change. Therefore, the relationship between subject's getting new attribute is implicit. By introducing this rule, we explicitly deal with this point by the changes in the corresponding subjects and attributes.

Here, we can refer to Simon (1996) where Simon stated: "[T]he behavior of the system at each level depended on only a very approximate, simplified, abstracted characterization of the system at the level next beneath" (Simon 1996, p. 16). Yoshikawa (1981) expressed a similar consideration as the second supposition of the general theory of the design process.

We then introduce a rule for the orthogonal representation of personal concerns. Our grammar contains definitions to express the following two views [see Fig. 6(a) and (b)]:

- Since E is a personal reason, I/we generate a design that will change F to G in the hierarchical representation of artifacts.

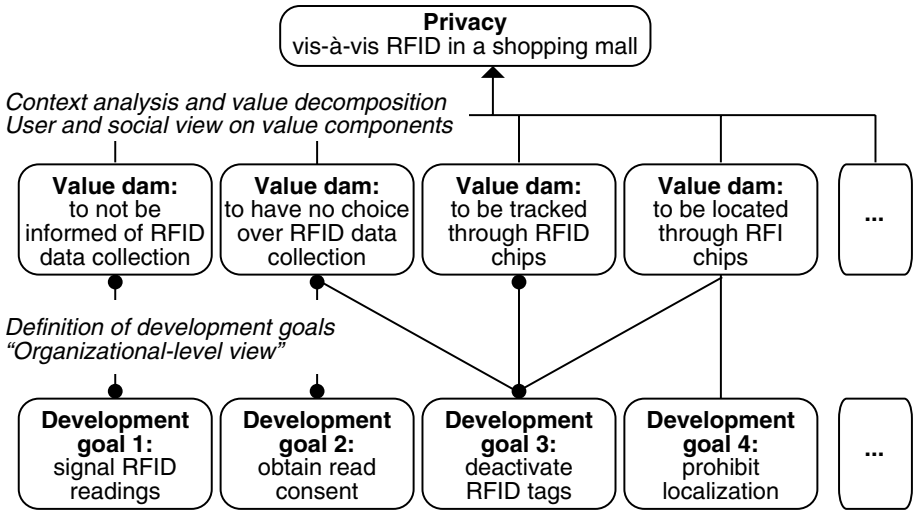

Fig. 7 Value dams for privacy in the exemplary RFID mall context (Spiekermann 2016)

- If F is changed to G in the hierarchical representation of artifacts, then H will change to J as effects on me/us.

3.3 Mathematical representation of design with discourse

To obtain a better understanding, the expressions in Figs. 5 and 6 can be represented by partial differential equations.[2] First, C in Fig. 5 has A as a component and is a function of time. Therefore, C is described as f_C in Eq. 1 and we can describe dC as a total differential equation:

$$C = f_C(t, A) \tag{1}$$

$$\mathrm{d}C = \frac{\partial f_C}{\partial t}\mathrm{d}t + \frac{\partial f_C}{\partial A}\mathrm{d}A \tag{2}$$

For the orthogonal representation, we can describe the function f_F for F and f_H for H in Fig. 6 as follows:

$$\mathrm{d}F = \frac{\partial f_F}{\partial t}\mathrm{d}t + E\mathrm{d}t \tag{3}$$

$$\mathrm{d}H = \frac{\partial f_H}{\partial t}\mathrm{d}t + \frac{\partial f_H}{\partial F}\mathrm{d}F \tag{4}$$

$$= \frac{\partial f_H}{\partial t}\mathrm{d}t + \frac{\partial f_H}{\partial F}\left(\frac{\partial f_F}{\partial t}\mathrm{d}t + E\mathrm{d}t\right). \tag{5}$$

Describing ideas as a tree structure consisting of these changes corresponds to constructing multivariate partial differential equations and implementing them to solve the problem.

[2] For more details, see: K. Sekiguchi, Mathematical representation of the design, July 17, 2013 at http://www.ethics-level.com.

Here, considering the relationships between various levels corresponds to the comments by Simon or Yoshikawa in Sect. 3.2 (Simon 1996; Yoshikawa 1981). And, Yoshikawa (2008) also thought that a function such as a service can be described as a differential equation.

3.4 Related work in detail

In the previous section, we compared and positioned our research for related work having a hierarchical representation. In this section, we will clarify the position of this research by showing examples of differences from the hierarchical representation proposed in IT ethics. Here, we refer to an influential related study focused on value dams and flows (Miller et al. 2007), which can also be described as a hierarchy. A case related to privacy issues associated with radio frequency identification (RFID) is shown in Fig. 7 (Spiekermann 2016).

The tree of value dams and value flows represents the hierarchy to define "development goals", as shown in Fig. 7. The point is that the decomposition of the privacy issue on RFID to the "User and social view" in Fig. 7 is considered as a concretization for detailed contexts, so it differs from the engineering decomposition done from the upper level to the lower levels based on causality. Furthermore, the development goals are described as "Organizational-level view", hence they should be placed in subjective perspectives and this concerns the axis we have set as orthogonal. Therefore, the representations for artifacts and personal concerns were mixed in Fig. 7, and engineers can be confused with them.

These two representations can be solved only after we define our three-dimensional representation. Note that we can describe the same case by applying the ethical design theory, as shown in Fig. 8.

The described user and social issues in Fig. 7 can be placed around the interaction level, as shown by (a)–(d)

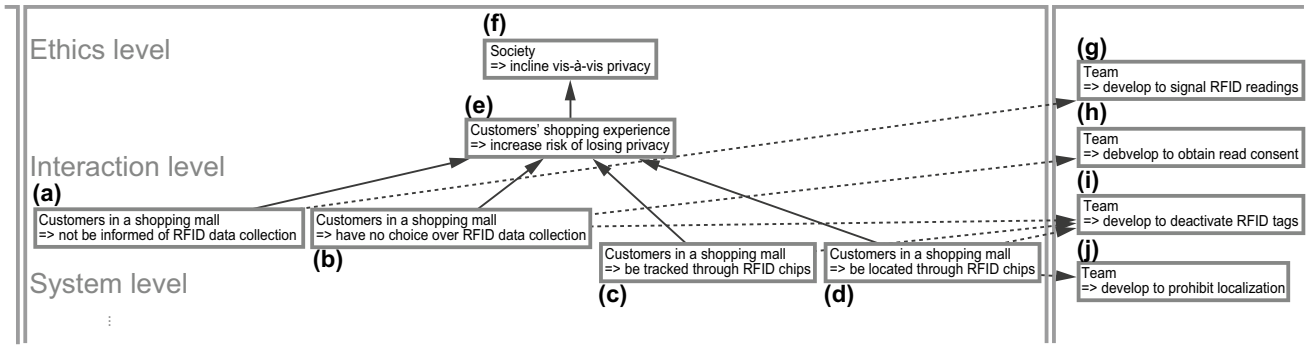

Fig. 8 The RFID case described using design with discourse whose descriptions are: (a) customers in a shopping mall ⇨ not be informed of RFID data collection, (b) Customers in a shopping mall ⇨ have no choice over RFID data collection, (c) Customers in a shopping mall ⇨ be tracked through RFID chips, (d) Customers in a shopping mall ⇨ be located through RFID chips, (e) Customers' shopping experience ⇨ increase risk of losing privacy, (f) Society ⇨ incline vis-à-vis privacy, (g) Team ⇨ develop to signal RFID readings, (h) Team ⇨ develop to obtain read consent, (i) Team ⇨ develop to deactivate RFID tags, (j) Team ⇨ develop to prohibit localization

in Fig. 8, because they relate to interactions. In this case, because the top description in Fig. 7 can be understood as a generic expression of the decomposed four descriptions, we could omit it because of the existence of (a)–(d) in Fig. 8. Note that we can add a description related to the shopping experience at the interaction level as (e) and related to the social values at the ethics level as (f) in Fig. 8. Finally, development goals are understood as future tasks to be placed as effects on me, such as (g)–(j), because they are generated from reflection on the artifacts by the development team. In this way, we can divide the problems in Fig. 7 into the hierarchical and orthogonal representation, as shown in Fig. 8, that align with the hierarchical perspective familiar to engineers.

Thanks to this difference, it can be clearly seen that the description method shown in Fig. 7 makes it difficult to elaborate upon further ideas and that our way of representation has solved this problem. For example, using Fig. 7, it is difficult to describe issues at higher levels such as (e)–(f) in Fig. 8, because such levels and the description methods that connect the changes between them are not clearly defined in Fig. 7, while our ethical design theory defined them. Furthermore, with respect to the lower levels, when a more detailed design such as subsystem, parts, materials of RFID system becomes necessary, it is difficult to describe it in Fig. 7 because the development goals already take space for such lower levels.

Although we can simultaneously set development goals at any of these levels, it is difficult to describe them using Fig. 7. For example, regarding (f) in Fig. 8, someone may consider how much a society protects and exposes privacy to be a matter of choice that further discussion on this problem is necessary in the future. It is also possible for someone to find challenges at the lower level that require the development of hardware materials that are hard to hack. These future tasks for each level could be clearly expressed for the first time by our three-dimensional representation; if we used Fig. 7, descriptions of artifacts' changes and future tasks would have become sandwiched and confused.

Moreover, there is no space to describe personal reasons why the designer considers such privacy issues and RFID technology needs to be improved. A designer might believe that privacy issues and RFID technology need to be improved because he/she suffered a privacy infringement in the past and feels strongly about solving the problem, or he/she believes that RFID is a key technology to solve privacy infringement and will become a big business. Our ethical design theory possesses the ability to describe personal reasons on the left side of the hierarchical representation. For example, justness of such personal reasons becomes important when we evaluate whether he/she is ethical enough as a professional engineer.

We then can describe the development goals in the hierarchical representation of artifacts as the next step design, but its explanation is omitted this time because of space constraint.

3.5 Rephrasing the gap between AI ethics and AI technology

Then, we can understand more clearly why practicing AI ethics is difficult for engineers. The first reason is because the ethics level is too high in the hierarchical representation of artifacts. Given that general engineers deal with the system level at the most, there exist exact gaps. Furthermore, if there is a large amount of information regarding the ethics level, this gap will also cause tremendous confusion.

Then, we can clearly understand that the propositions at the ethics level play the role of Wittgenstein's hinge in many engineering designs. Wittgenstein argued:

> That is to say, the questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn.
> That is to say, it belongs to the logic of our scientific investigations that certain things are indeed not doubted.
> But it isn't that the situation is like this: We just can't investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinge must stay put (Wittgenstein 1969).

Therefore, we did not tend to doubt propositions at highest levels such as the ethics level for practicing standard design, but will be able to utilize such reconsideration for seeking alternative designs.

In addition, it is also confusing that investigations concerning the ethics level and personal concerns are mixed in AI ethics. Our perspectives separate them clearly by orthogonalizing them. Therefore, these investigations can be easily handled by engineers using their familiar hierarchical perspective.

Our aim can, therefore, be rephrased. We aim to support engineers to clearly understand and design from the ethics level when required. Then, although we believe that our ethical theory shows the truth, we will evaluate from the viewpoint of usefulness what we can do with this theory.

4 Overview of the tool for use with knowledge base of AI ethics called dfrome

To support the application of the ethical design theory to AI engineering, we implemented an organic and dynamic tool called dfrome. In this section, we introduce the overview of dfrome.

4.1 Construction of dfrome

As is shown in Fig. 9, dfrome utilizes three technologies: an editor, a cloud environment, and an investigation engine.

Here, the editor provides an ethical perspective on the extension of engineering and decreases the editing load to allow the user to immerse into design activity. The cloud environment where database (knowledge base) is stored omits the efforts of the user's installation of the application, and makes it possible to manage the design ideas independently of the local environment. This environment also makes it easy for users to share ideas. The investigation engine has the function of calculating the distance among AI ethics and the design ideas and calculating scenario paths to be considered in the context of the user.

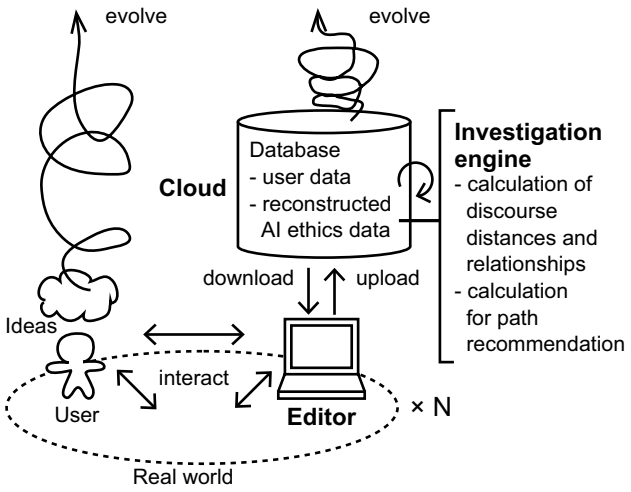

Fig. 9 Overview of the organic and dynamic tool for use with knowledge base of AI ethics called dfrome

The contents of the knowledge base of AI ethics contain several principles and cases of AI ethics. Our tool also contains other descriptions, such as the methods of machine learning, the original design ideas of the authors, and so on.

Moreover, as shown in Fig. 9, the idea possessed by the user and the contents stored in the knowledge base are supposed to evolve dynamically by the interactions between the user and the tool in the real world.

4.2 Expected use case

Here, we describe an expected use case of dfrome. For example, we suppose a designer is designing an AI self-checkout system. By referring to the documents discussed in AI ethics, we can see that this AI system can deprive human persons of their jobs (IEEE 2016, 2017). Furthermore, up to the ethical level, the designer can consider that there is the probability that this AI system will decrease people's economic power and thereby their ability to live a life of freedom and equality. Therefore, the designer would know that this AI system needs to be updated so that it does not negatively influence people's lives.

Then, it may be possible to obtain alternatives by referring to best practices. For example, we can consider an AI system that supports cashiers of a bakery shop. This AI system would ease the heavy task of the cashier by automatically recognizing the kind of breads whose lineup partially changes everyday and by providing a list of candidates when the recognition result is not sure (Oka and Morimoto 2015, 2016). The point is that, to reduce the workload of cashiers, the AI does not completely replace the human jobs but

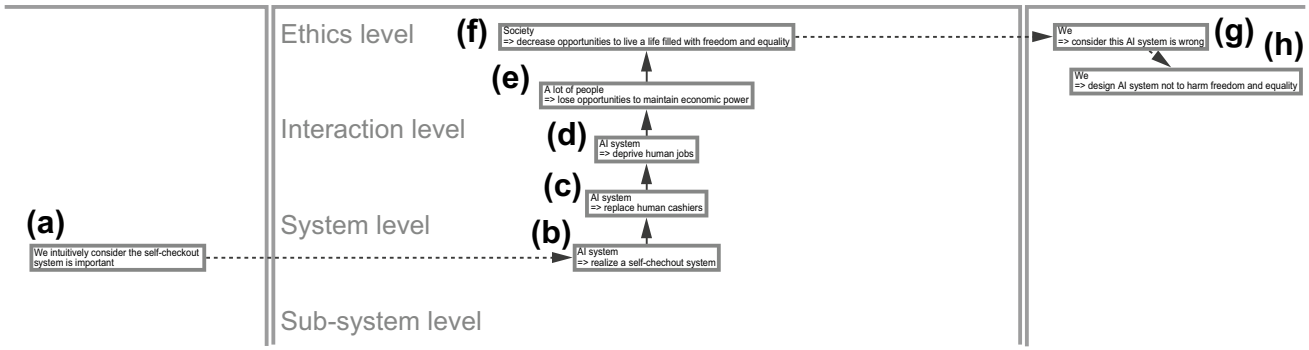

Fig. 10 Initial design of AI system for supporting shop cashiers whose descriptions are: (a) ∵ we intuitively consider the self-checkout system is important, (b) AI system ⇨ realize a self-checkout system, (c) AI system ⇨ replace human cashiers, (d) AI system ⇨ deprive human jobs, (e) A lot of people ⇨ lose opportunities to maintain economic power, (f) Society ⇨ decrease opportunities to live a life filled with freedom and equality, (g) We ⇨ consider this AI system is wrong, (h) We ⇨ design AI system not to harm freedom and equality IEEE (2016), IEEE (2017)

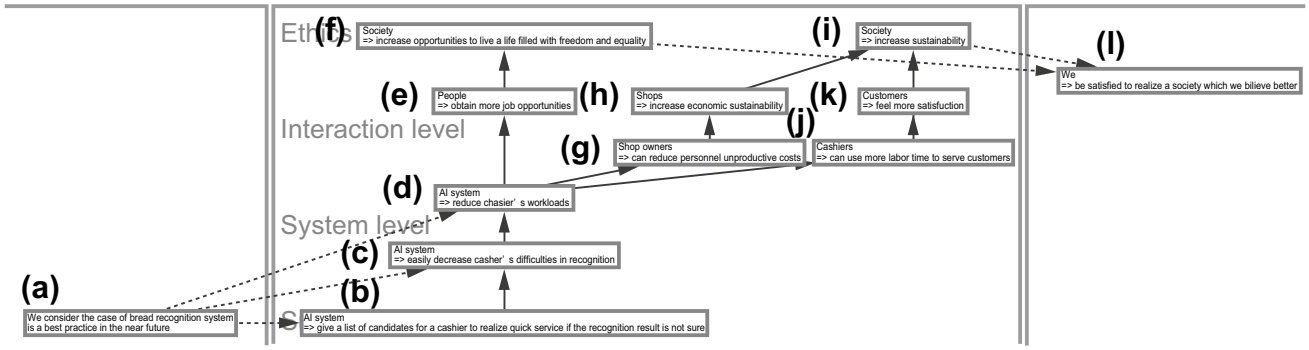

Fig. 11 Revised version of design of AI system for supporting shop cashiers whose descriptions are: (a) ∵ we consider the case of bread recognition system is a best practice in the near future, (b) AI system ⇨ give a list of candidates for a cashier to realize quick service if the recognition result is not sure, (c) AI system ⇨ easily decrease cashiers' difficulties in recognition, (d) AI system ⇨ reduce cashier's workloads, (e) People ⇨ obtain more job opportunities, (f) Society ⇨ increase opportunities to live a life filled with freedom and equality, (g) Shop owners ⇨ can reduce personnel unproductive costs, (h) Shops ⇨ increase economic sustainability, (i) Society ⇨ increase sustainability, (j) Cashiers ⇨ can use more labor time to serve customers, (k) Customers ⇨ feel more satisfaction, (l) We ⇨ be satisfied to realize a society which we believe better (IEEE 2016, 2017; Oka and Morimoto 2015, 2016)

helps persons achieve better performance by enabling the human–machine interactions.

Then, this AI system can expand job opportunities rather than reduce them. Furthermore, we can expect that shop owners can reduce the cost of unproductive personnel, and cashiers would have more time to serve customers. Therefore, this can also make society more sustainable at the ethics level.

In this way, rather than merely considering ethical considerations as a constraint for creative activities, these considerations can be used as a means of enhancing better design. An overview of the initial design is shown as Fig. 10, and one of revised versions is shown in Fig. 11. Dfrome is designed to promote engineers' practice of such design.

5 Knowledge liquidization and crystallization model

Knowledge management in dfrome is based on a dynamic model called knowledge liquidization and crystallization (Hori 2004). Our aim is also not necessarily to provide precise information but to suggest a viewpoint that can be a trigger for further awareness to create a better design. From the next subsection, we introduce two types of knowledge liquidization and crystallization.

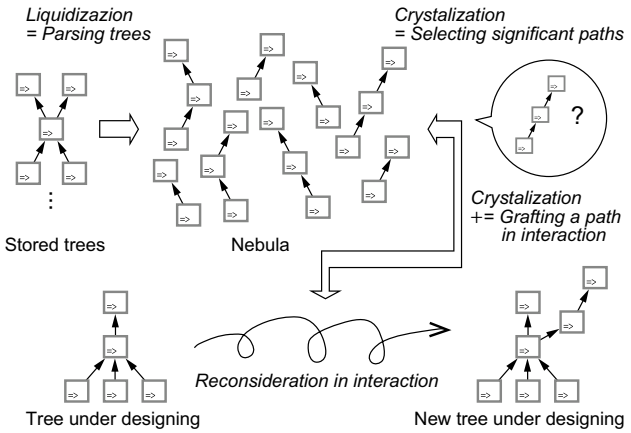

Fig. 12 Overview of the processing of path recommendations

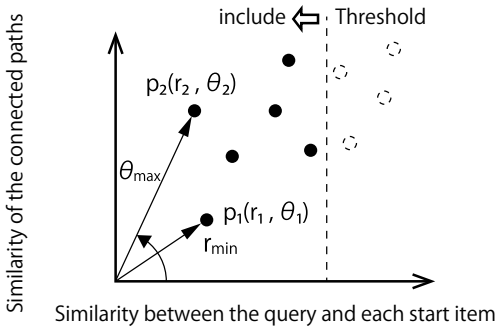

Fig. 13 Geometrical representation of path distances

5.1 Calculation of discourse distances

With dfrome, users can calculate the distance between their own ideas and the already-stored trees. Here, liquidization corresponds to the way of storing AI ethics and design ideas to be managed in units of trees. It does not involve fixing the whole database as a statistic or saving AI ethics in units of documents; this allows designers to dynamically calculate mutual distances using their own ideas that can be updated each time as a query. Crystallization corresponds to the process of presenting discourses by computing the distance using the designer's own idea as a query. In addition, each time a new idea is registered, the candidates for presentation increase, and the corpus becomes ready to be updated at the next daily update process. The knowledge base evolves in this way.

By checking similar trees, the user can confirm whether their own ideas have already been proposed, or the user can see the ideas, such as AI ethics, that have already been counted. Furthermore, by checking the furthest trees, users can check the viewpoints that they have missed.

Technically speaking, we calculate a vector space by applying Doc2Vec (Řehůřek and Sojka 2010) because it allows us to respond flexibly to synonyms and data granularity compared with bag-of-words and term frequency–inverse document frequency (tf–idf). We then choose to use the distributed bag-of-words (DBOW) model that is simpler than the distributed memory (DM) model because of the better ability of DBOW to manage small datasets. DBOW can also avoid the random processes that DM has and the realization of quick responses. Then, the paragraph vector of each tree can be calculated based on the terms contained. With the two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, the similarity equation is given in Eq. 6.

$$\text{similarity} = 1 - \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|_2 \|\boldsymbol{v}\|_2}. \tag{6}$$

When a tree is sent as a query that is described by the editor but not necessarily saved in the database, the similarities between this and all the active (not deleted) and public trees will be calculated. An example of this is shown in Sect. 6.2.

5.2 Calculation for path recommendation

The path recommendation adopts more explicitly the knowledge model on dfrome as shown in Fig. 12.

Here, the designer receives a recommendation of a similar path or a divergent one from the accumulated paths using an item (a description of a change) in the tree as a query. Then, the design ideas will be updated through such interactions. When these new design ideas are saved, the database will also be updated and the candidate group of paths will be updated, and so on.

In detail, we calculate the similarities between paths using the geometry in Fig. 13.

The horizontal axis in Fig. 13 shows the similarity between the query item of the designer's tree and the start item of the comparison objects. The vertical axis shows similarity between the connected paths that extend from the query or the start items. Each plot signifies a candidate path. To recommend such paths, we set two metrics. One is the absolute value of similarity between the paths, and the other is an angle as shown in Fig. 13.

Then, the calculation function for a similar path recommendation is based on the absolute value and is given as follows:

$$\text{totalSim} = \sqrt{\text{queryItemSim}^2 + \text{connectedPathSim}^2} \tag{7}$$

$$\sim \text{queryItemSim}^2 + \text{connectedPathSim}^2. \tag{8}$$

Here, the term Sim is an abbreviation for similarity. The term queryItemSim refers to the similarity between the query item and start item, and the connectedPathSim shows the similarity between the connected paths.

However, the total similarity for the divergent path recommendation is based on the angle and is shown in Eq. 10.

$$\text{totalSim} = -\arctan\left(\frac{\text{connectedPathSim}}{\text{queryItemSim}}\right) \tag{9}$$

$$\sim -\frac{(1+\text{connectedPathSim})}{(1+\text{queryItemSim})}. \tag{10}$$

Here, we define the total similarity as a negative value corresponding to the relationship between the magnitudes in Eq. 8. We then add one to the denominator to correspond to the case in which the denominator becomes zero. The numerator corresponds to the denominator.

Finally, we rank the paths according to the above-mentioned evaluation functions. Here, it is probable that similar paths are duplicated in the result: therefore, we apply a clustering method to solve the problem.

First, we preprocess each path to obtain a list of ids of similar paths whose similarity in Eq. 6 is smaller than the threshold (currently it is $1-\cos 30°$). Then, by considering the paths described in the similar paths' list, a path that has not already existed in the results will be added to the results one by one in order.

These are the basic ideas; there are several exceptional processes, but their explanations are omitted this time because of space constraint. Examples of path recommendation are shown in Sect. 6.3.

5.3 Related work in detail

We have already discussed the comparison of the hierarchical representations. Next, we will compare dfrome's way of describing ideas based on the changes in the related field. First of all, we compare our calculation method with a well-known method called knowledge graphs. Knowledge graphs are to "model information in the form of entities and relationships between them", and some of the models deal with latent features (Nickel et al. 2016) and the path query (Guu et al. 2015).

Our calculations owe their network form to the ethical design theory, which gives rise to the following differences. First, our method describes the changes in the attributes of entities, not the entities themselves; therefore, our representation became a form of partial derivative. Then, we can find several latent features in our method, too. For example, the values at higher levels have preconditions, such as social backgrounds; the personal dimensions correspond to the context of each subject. Finally, our purpose is to provide creativity support; therefore, extracting paths for diverging users' thinking became more important than extracting the ordinary facts.

Table 1 Overview of stored active and public trees (IEEE 2016, 2017; AI Network 2017a, b, c; FLT 2017; Oka and Morimoto 2015, 2016; Nickel et al. 2016; Le and Mikolov 2014; Schmid 2010; Krizhevsky et al. 2012)

Tag	Document	Details	*N*
EAD	Ethically aligned design	Principles of version 1	4
EAD2	Ethically aligned design	Principles of version 2	5
AAP	Asilomar AI principles	Ethics and values	14
RDP	AI R&D guidelines	Principles	10
RDU	AI R&D guidelines	Cases	10
BP	Best practices papers	Best practices	5
ML	Machine learning papers	Overview of the algorithm	5
OTH	Others	Authors' design, etc.	6
			59

Table 2 Overview of additionally stored active and public trees (Yanmaz et al. 2017; Terada 2018; Japan 1946)

Tag	Document	Details	*N*
DRN	Drone papers or cases	Overview of the idea	6
SAT	Nano or micro-satellite papers	Overview of the idea	2
OTH	Others	Authors' design, etc.	2
			10

Given these differences, this research provides a new perspective to the knowledge graph as well. For example, this study provides a novel theme of obtaining a hierarchical representation by calculating the partial differentiation of the knowledge graph.

The other related work is Neuron Data Inc. (1996) where the reasons for such changes are dealt with as three-dimensional representations, although we express them in a two-dimensional hierarchy by describing a change such as Fig. 5: we use the third dimension for a different purpose, for describing personal concerns. Our way of describing changes is more intuitive for describing design ideas, because it more clearly corresponds to the design thinking to changes that occur from the parameter level to the target level.

In addition, we can understand that the networks of Eqs. 2 and 5 are similar to the neural networks because both of them are based on the chain rules. Utilizing neural networks is one of our future works, too.

6 Cases for discussion

To confirm the expected effects, we introduce concrete cases for discussion. The overview of data used this time is shown in Table 1. (As of March 30th, 2018, there exist 59 trees that can be used for the experimental processing).

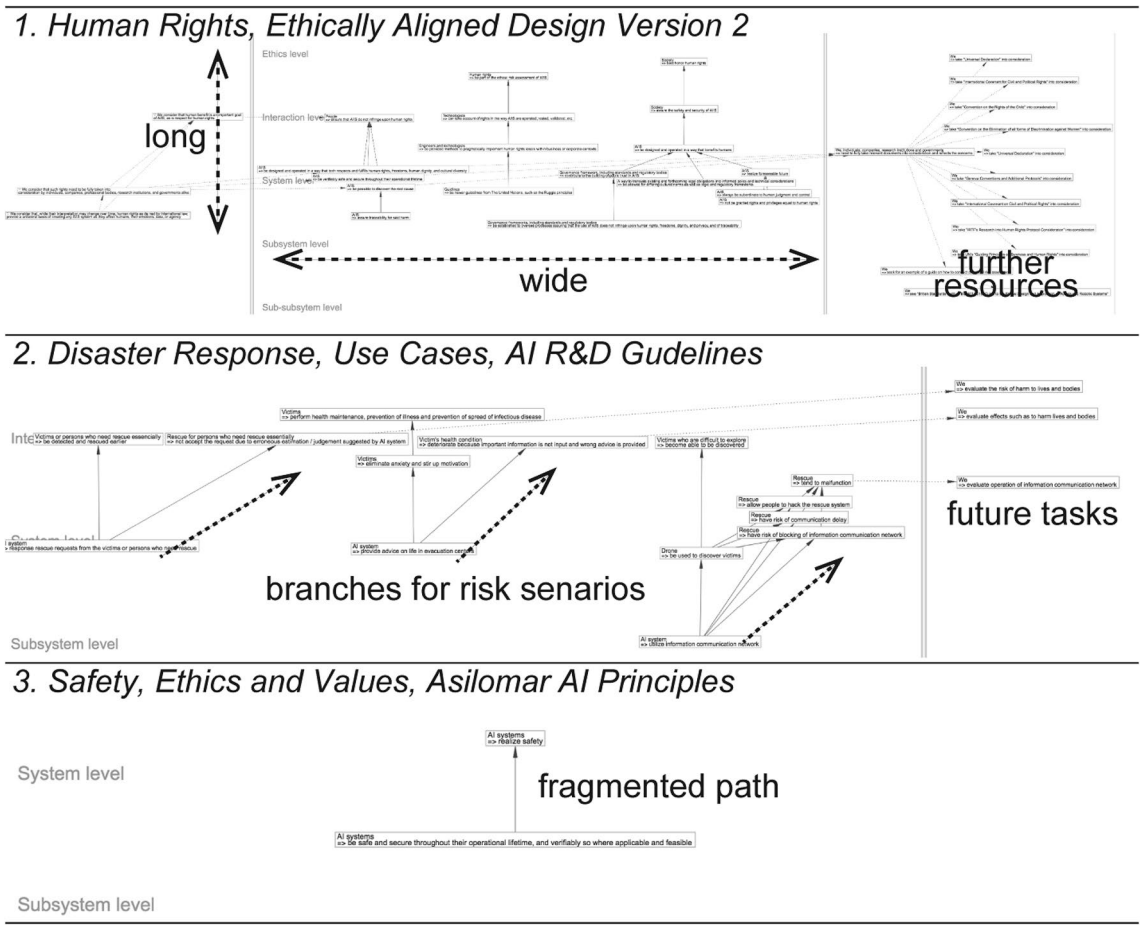

Fig. 14 Example of clarification of structural differences (IEEE 2017; AI Network 2017c; FLT 2017)

Here, we introduce three cases. The first case involves the visualization of differences in the structure of AI ethics so that the ethics can be clearly understood at a glance in a standardized way. In the second case, the designer can compute and list the distance between the idea at hand and the accumulated ideas. The third case involves receiving the recommendation of the scenario paths that should be considered by the designer.

For the third case, we first executed three examples with using the data in Table 1: to find a path to the ethics level, a path of risk scenarios and a path of best practices. Then, we also did an additional experiment for further discussion using the latest version of the trees in Table 1 and ten new trees which referred to descriptions such as Yanmaz et al. (2017), Terada (2018), Japan (1946) shown in Table 2 to introduce an example of discussion with humanities, social sciences, and fine arts experts and a case of a self-application of dfrome (as of September 5th, 2018, there exist 69 = 59 + 10 trees that can be used for experimental processing).

In both experiments, the execution program was based on the latest version deployed to the production environment at the time of data acquisition. We introduce the experimental results in this section and the discussion in the next section.

6.1 Clarifying differences of structures among AI ethics and design ideas

We confirmed how each AI ethics differs in its structure, because the differences in its structure was visualized in the standard way of representation. Figure 14 shows three typical examples.

In the first case, the trees of the ethically aligned design (IEEE 2017) tend to become wide and long because they enumerate many issues. Furthermore, because this report introduces many further resources, it is possible to describe the future tasks for effects on me. Then, the trees of AI Network (2017c) tend to have upward branches, because they deal with risk scenarios with future tasks. Finally, the trees of FLT (2017) become fragmented, because the principles provide wide viewpoints in the form of a list of short sentences.

Table 3 Results of tree distance calculation (IEEE 2016, 2017; AI Network 2017b, a, c; FLT 2017; Nickel et al. 2016; Le and Mikolov 2014; Schmid 2010; Krizhevsky et al. 2012; Oka and Morimoto 2015, 2016)

Rank	Cases	
	Querying the tree of Human rights, EAD2	Querying the tree of Paragraph vectors, ML
1	Human rights, EAD2	Paragraph vectors, ML
2	Human benefit, EAD	Convolutional neural network, ML
3	Human values, AAP	Education and human resource development, RDU
4	Ethics, RDP	Bread recognition system, BP
5	Safety, AAP	Knowledge graph, ML
55	Manufacturing and maintenance, RDU	Human control, AAP
56	Accountability, RDP	Health, RDU
57	Dynamic mode decomposition, ML	Disaster response, RDU
58	Convolutional neural network, ML	Value alignment, RDP
59	AI ethics library, OTH	Responsibility, AAP

The same effects can be true for the description of design ideas.

6.2 Presenting distances among AI ethics and design ideas

Next, we investigated the case for calculating the distance among AI ethics and design ideas. This time, we queried using human rights of ethically aligned design (IEEE 2017) and paragraph vectors of machine learning (Le and Mikolov 2014), as shown in Table 3.

These examples suggest that it is possible to support users based on semantic relationships because they could calculate ideas of similar subjects or ideas in the categories close to one another. For example, human rights was calculated to be close to the trees of other general ethics. Also, paragraph vectors was calculated to be close to the machine learning methods, especially the convolutional neural network that is also a kind of neural networks.

However, regarding the trees calculated as distant ones with lowest ranks, there is no clear trend. Therefore, further consideration is necessary for each result.

6.3 Recommending scenario paths

Finally, we introduce three examples of path recommendation: the first is an example of the recommendation for considering ethical scenarios. The second is one for considering risk scenarios. And, the third is one for considering best practices. For this time, the tree of paragraph vectors (Le and Mikolov 2014) is set for a sample design idea for the first three examples, and the tree of drone networks (Yanmaz et al. 2017) for the example of discussion with humanities, social sciences, and fine arts experts, and the tree of dfrome itself for the example of self-application.

6.3.1 Path to the ethics level

We introduced a case to check the scenarios connected to the ethics level (see results in Fig. 15).

At first, we could obtain (A) in Fig. 15 for divergent thinking by querying the item (a): Paragraph vectors is changed to provide state-of-art results on several text classifications (Le and Mikolov 2014). The path (A) suggested that it can foster utmost use of AI systems through social education. In addition, we could obtain the path (B), which was similar to and connected to (e); this suggests that society will change to emphasize human rights by assuring safety and security.

Therefore, in this case, we can finally notice that paragraph vectors relates to human rights at the ethics level.

6.3.2 Path of risk scenarios

The second example is checking risk scenarios as shown in Fig. 16.

Here, we could obtain the path (A) in Fig. 16 for divergent thinking which suggested the probability of people's privacy being infringed upon (see (b), Fig. 16) as a risk scenario for paragraph vectors. Furthermore, due to (c) in effects on me in Fig. 16, we can confirm that someone considered that we should evaluate the effect of privacy infringement.

6.3.3 Path of best practices

Finally, we provide an example for the best practices. Here, there are two probabilities. One probability is reconstructing the already registered scenarios of best practices in the current context. The other probability is dynamically

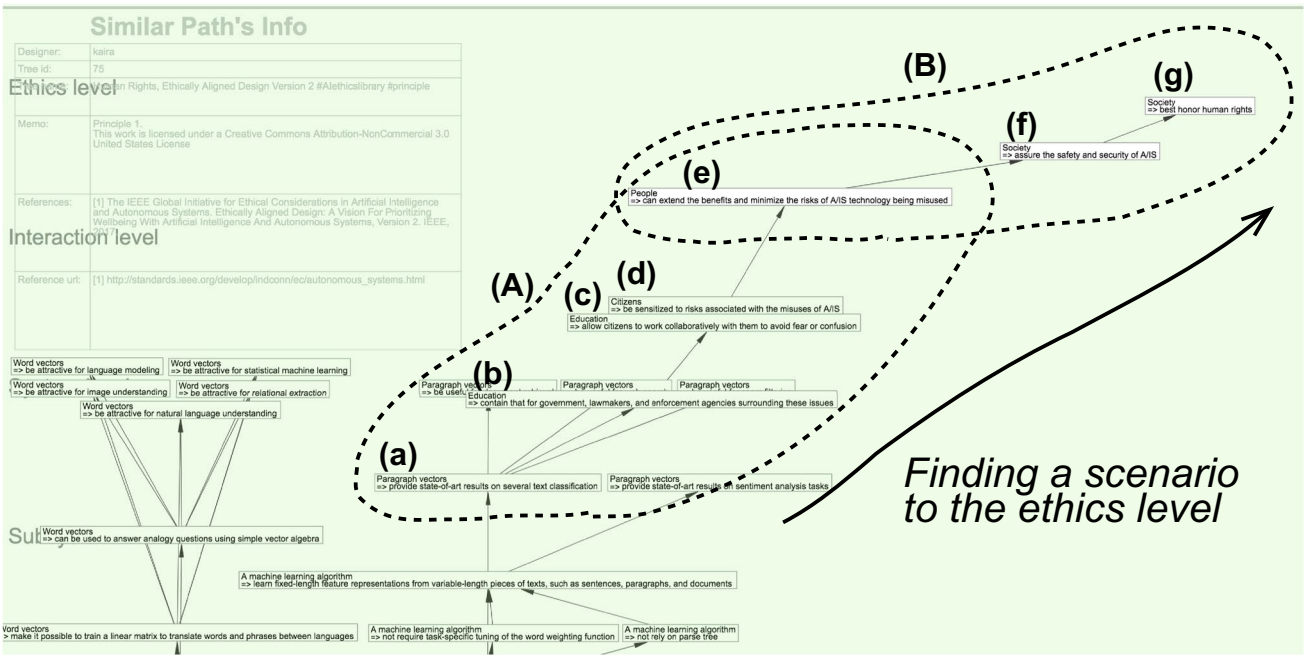

Fig. 15 A case of the path recommendation to the ethics level whose descriptions are: (a) Paragraph vectors ⇨ provide state-of-art results on several text classification; (b) Education ⇨ contain that for government, lawmakers, and enforcement agencies surrounding these issues; (c) Education ⇨ allow citizens to work collaboratively with them to avoid fear or confusion; (d) Citizens ⇨ be sensitized to risks associated with the misuses of A/IS; (e) People ⇨ can extend the benefits and minimize the risks of A/IS technology being misuses; (f) Society ⇨ assure the safety and security of A/IS; (g) Society ⇨ best honor human rights Le and Mikolov (2014), IEEE (2017; A)'s rank is 2; (B)'s rank is 7

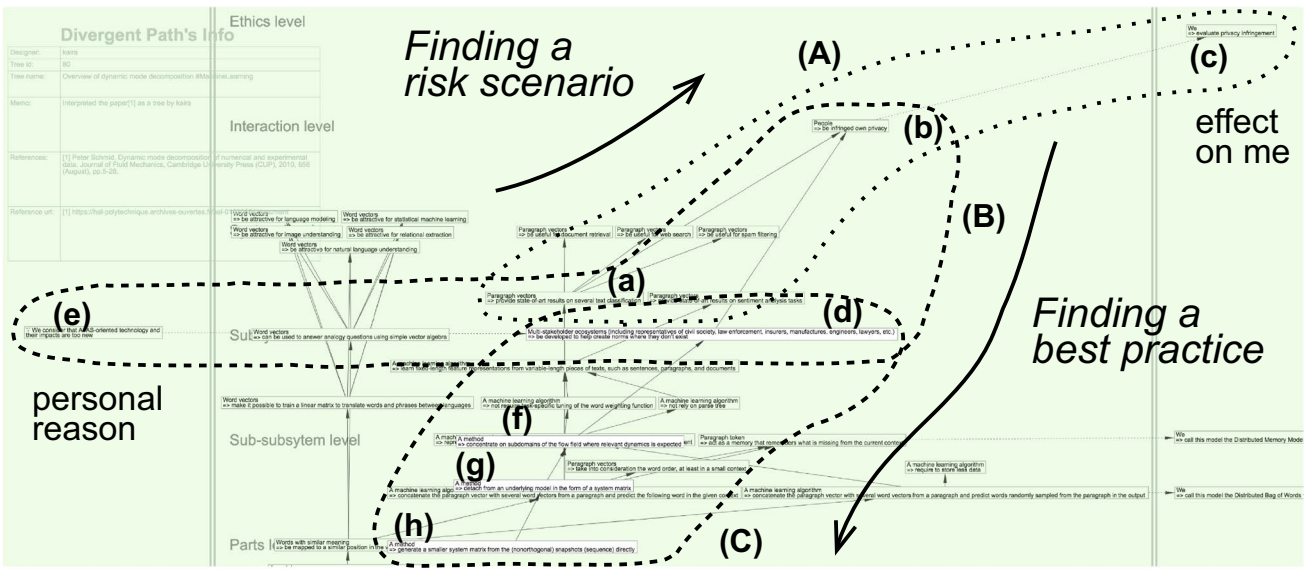

Fig. 16 A case of the path recommendation for finding risk scenarios and best practices whose descriptions are: (a) Paragraph vectors ⇨ provide state-of-art results on several text classification; (b) People ⇨ be infringed own privacy; (c) We ⇨ evaluate privacy infringement; (d) Multi-stakeholder ecosystems (including representatives of civil society, law enforcement, insurers, manufactures, engineers, lawyers, etc. ⇨ be developed to help create norms where they don't exist; (e) ∵ we consider that AI/AS-oriented technology and their impacts are too new; (f) A method ⇨ concentrate on subdomains on the flow field where relevant dynamics is expected; (g) A method ⇨ detach from an underlying model in the form of a system matrix (h) A method ⇨ generate a smaller system matrix from the (nonorthogonal) snapshots (sequence) directly (Le and Mikolov 2014; AI Network 2017c; IEEE 2016; Schmid 2010; A)'s rank is 2, (B)'s rank is 133 (total results are 135); (C)'s rank is 20

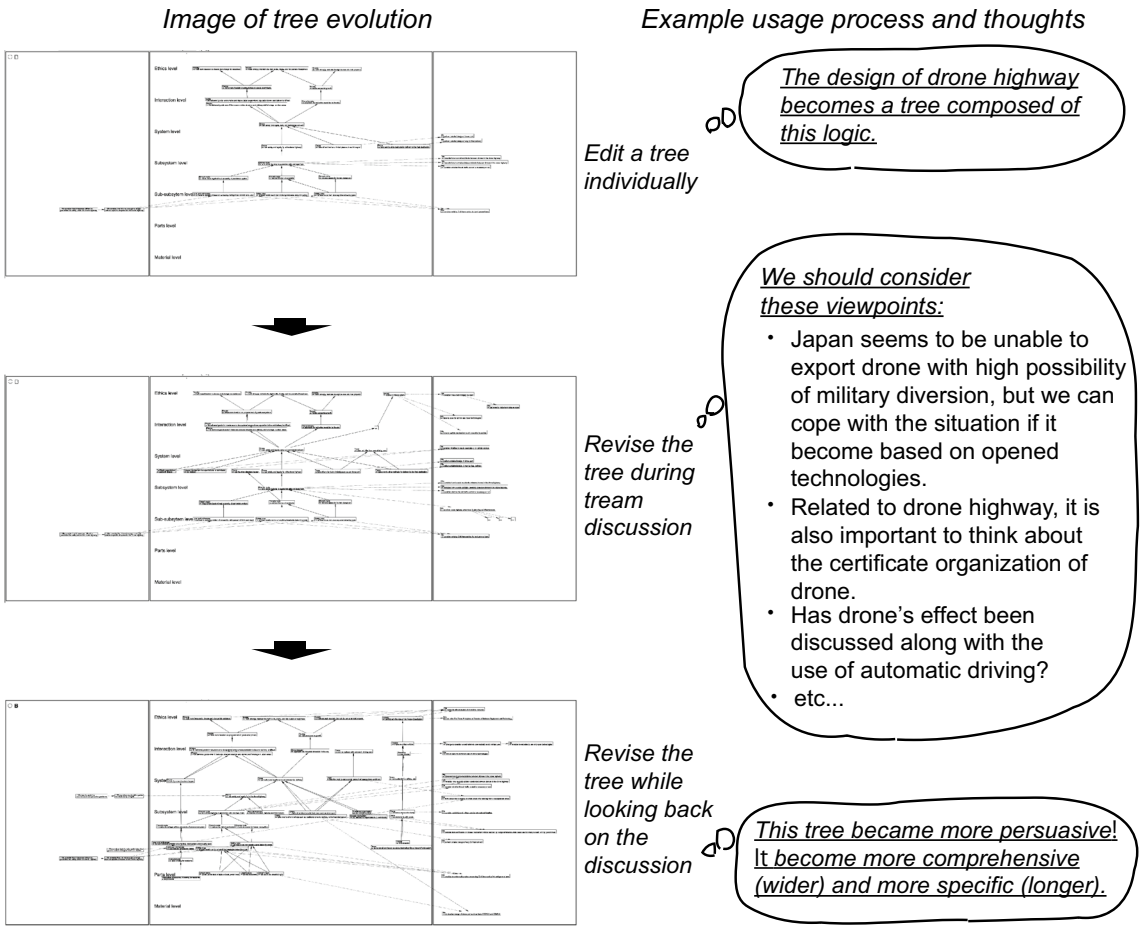

Fig. 17 An image of the process of discussing drone delivery and drone highway with humanities, social sciences, and fine arts experts (Terada 2018; Japan 1946)

constructing the scenarios by acquiring already registered technical paths, such as machine learning.

In Fig. 16, to consider the countermeasures against (b), we could be recommended the path (B) that suggested a multi-stakeholder ecosystem for completing the norms as (d). Here, we could find the recommended path by referring to the path whose similarity is calculated to be far from (b), which is an inverse proposition of the initial path; this is not necessarily true but is useful for further consideration. The item (b) must be duplicated to express the denial case later. Finally, we could obtain the dynamic mode decomposition by path (C) as an extension of the idea for developing the ecosystem.

In this way, we could find a novel solution, for example, to investigate the dynamic mode decomposition for creating complemental norms by investigating the dynamics of multi-stakeholder ecosystem.

6.3.4 Example of discussion with humanities, social sciences, and fine arts experts

Our tool also aims to enhance the collaboration between experts of different sectors; we are already discussing our study with humanities, social sciences, and fine arts experts to realize collaboration.

For example, we discussed a case of drone delivery and drone highway with an AI ethics/AI expert and a public law expert. The discussion process of the drone delivery and drone highway is shown in Fig. 17. To prepare for the discussion, one of the authors individually edited a tree of the idea of the drone delivery and drone highway by referring to a related work (Terada 2018). Then, the tree was presented and edited in the discussion and updated by the author after the discussion.

We have discussed what kind of ethical values should be set from the ethics level. And we have confirmed that rights stated in the Constitution of Japan, e.g., right to life, liberty, and pursuit of happiness (Japan 1946), can provide foundations of such ethical values when we design legally

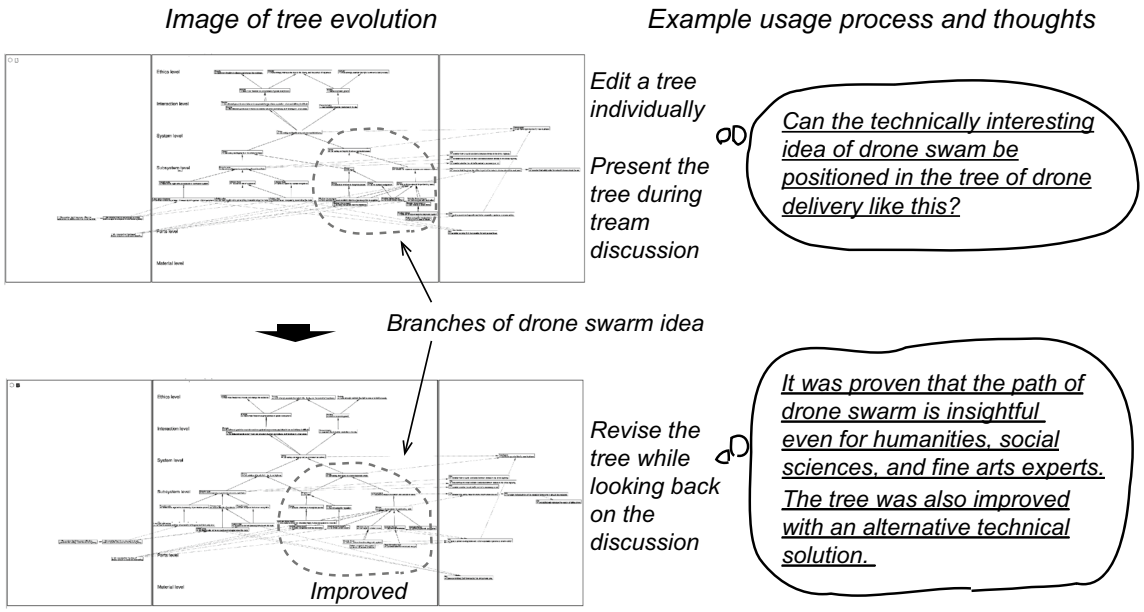

Fig. 18 An image of the process of discussing drone swarms with humanities, social sciences, and fine arts experts (Terada 2018; Japan 1946)

considered system, because each law is designed basically not to conflict with the constitution. Theoretically, it also becomes possible to doubt from this level such as to add an entirely novel value.

Furthermore, we discussed how our theory and dfrome could deal with the "dual use" case. Then, we confirmed a procedure to deal with such a case as follows. Dual use scenarios can be visualized by having upward branches. These branches will make it clearer to discuss the validity of the highest objectives, e.g, whether the dual should be allowed or controlled. Then, if designers decided to control the dual, they can realize it by modifying the design at the lower level, so that the above-mentioned upward branch will not appear.

For example, it was discussed that certain types of drone tend to have a high possibility of military diversion. Therefore, it cannot be exported overseas because it conflicts with the Peace Constitution of Japan (Japan 1946). However, it was suggested that if we could construct the drone system from already opened technologies, there would be no technical advantage against the opposite party. Thus, using transparent technologies may be one of the means of avoiding dual use scenarios.

We also discussed the usefulness of dfrome in finding alternative design solutions, e.g., drone swarm for safe and legal drone delivery. The discussion process of drone swarm is shown in Fig. 18. We confirmed that the technically interesting idea of "drone swarm" which has the potential to realize high robustness and reliability (Jassowski and Thirunahari 2018) can be visually designed with dfrome. And, we found that the idea of drone swarm is also insightful from the perspective of public law expert, because present laws and their application do not consider it (Ministry of Land, Infrastructure, Transport and Tourism 2015). As Fig. 18 shows, we were able to confirm this point visually as a branch and improve the tree with a new technical solution mentioned in the discussion. (We also intent to present another research paper extending these results.)

This kind of consideration can be easily practiced even with our current dfrome. For example, Fig. 19 is the result of receiving an upward scenario recommendation of dual use for a sample tree of drone networks (Yanmaz et al. 2017). The query is the description of the item "Small multicopters is changed to advance in technology and commercially available vehicle". Then, it was suggested that the society may be inclined to military applications.

As considered above, it can be expected that such a branch will not be connected by making small drones based on transparent technologies. Exactly, by appending such a description to the query, i.e., "Small multicopters is changed to advance in technology and commercially available vehicle *based on transparent technologies*" with the former query, the appearance of a military path was pushed back from the sixth rank result to the sixteenth rank.

This way, the discussion of dual use becomes clearer using ethical design theory and dfrome which provide procedures for reconsidering an engineer's own ideas. Although the results of dfrome are not absolute and its results guarantee nothing, it can provide designers with the awareness to reconsider their design ideas.

Therefore, high-quality contents can be registered to dfrome through bidirectional collaboration with humanities, social sciences, and fine arts experts. Dfrome can provide

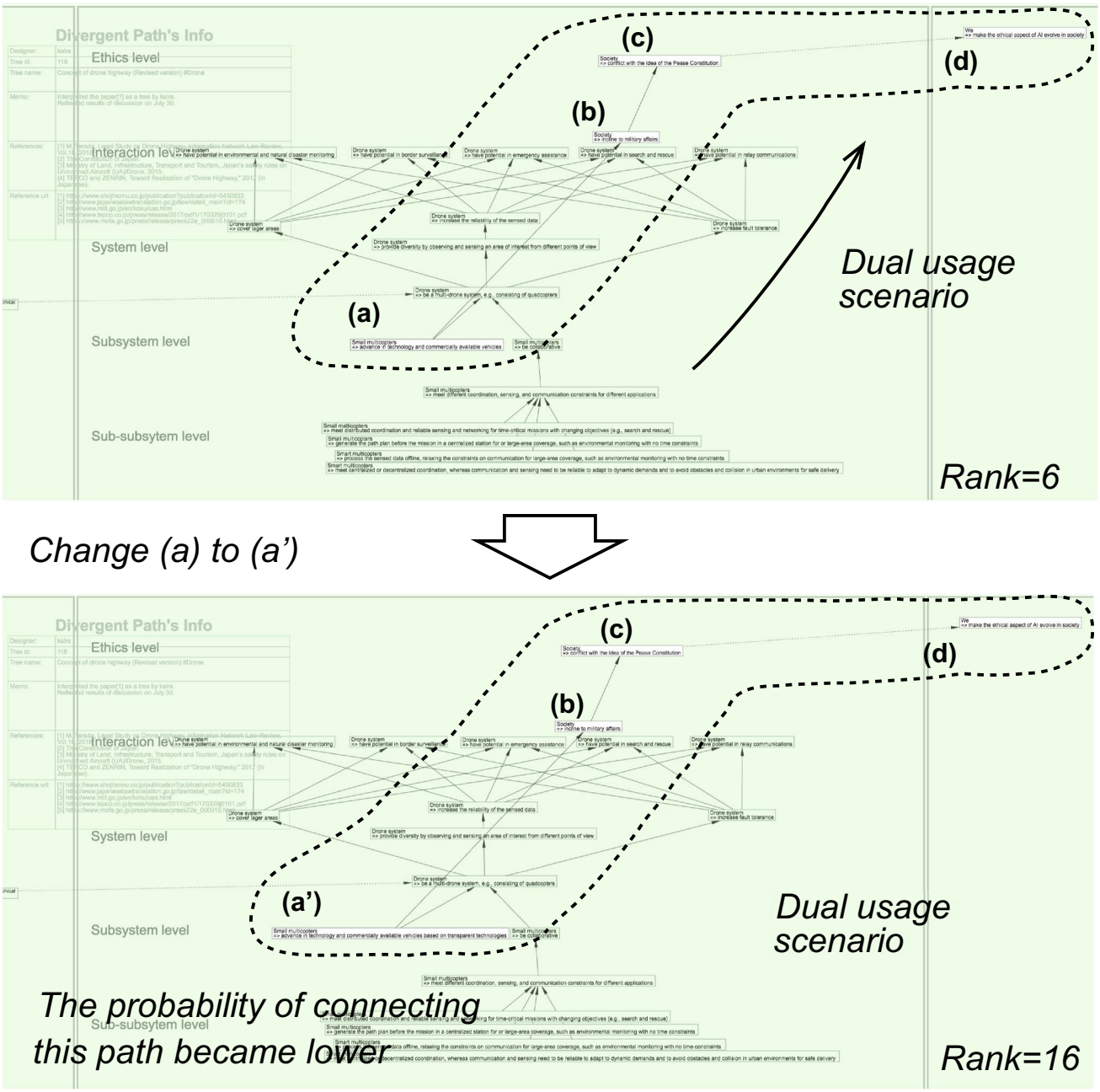

Fig. 19 A case of the path recommendation for finding dual use scenarios whose descriptions are: (a) Small multicopters ⇨ advance in technology and commercially available vehicles; (b) Society ⇨ incline to military affairs; (c) Society ⇨ conflict with the idea of the Peace Constitution; (d) We ⇨ make the ethical aspect of AI evolve in society; (a') Small multicopters ⇨ advance in technology and commercially available vehicles based on transparent technologies (Yanmaz et al. 2017; Terada 2018; Japan 1946); rank is from 6 to 16

the necessary contents and procedures to utilize these contents for all users. Furthermore, this case became an example of how to practice meta-ethics, e.g., how to set values at the ethics level through defining what "good" means for each design. The example in Sect. 6.3.3 showed how users can use dfrome to find best practices which may be useful in coming up with alternative solutions.

6.3.5 Example of self-application

Finally, the results of the self-applying case of dfrome are described in this section.

First, the idea of dfrome can be described in the form of a tree (Fig. 20). Since the description of the tree has been updated frequently, we used the version that is consistent with the description in this paper. As shown in this tree, it is

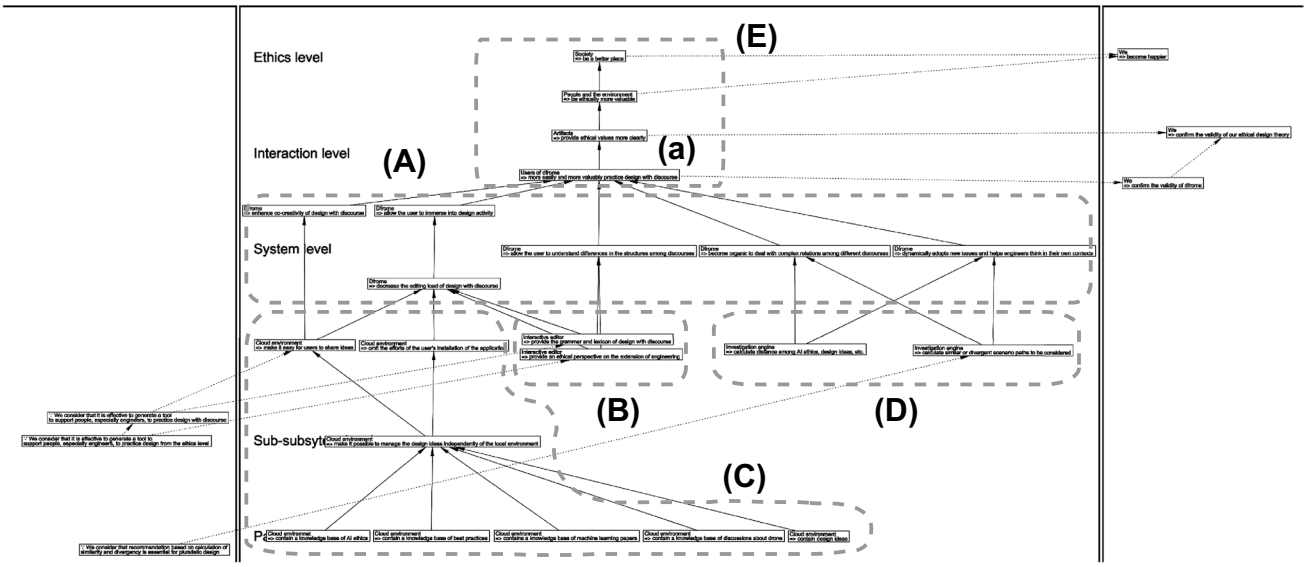

Fig. 20 An overview of a tree of dfrome's self-application; (a) Users of dfrome ⇨ more easily and more valuably practice design with discourse; (A) Functions of dfrome around the system level; (B) Interactive editor; (C) Knowledge base in a cloud environment; (D) Investigation engine; (E) Scenario to make society a better place

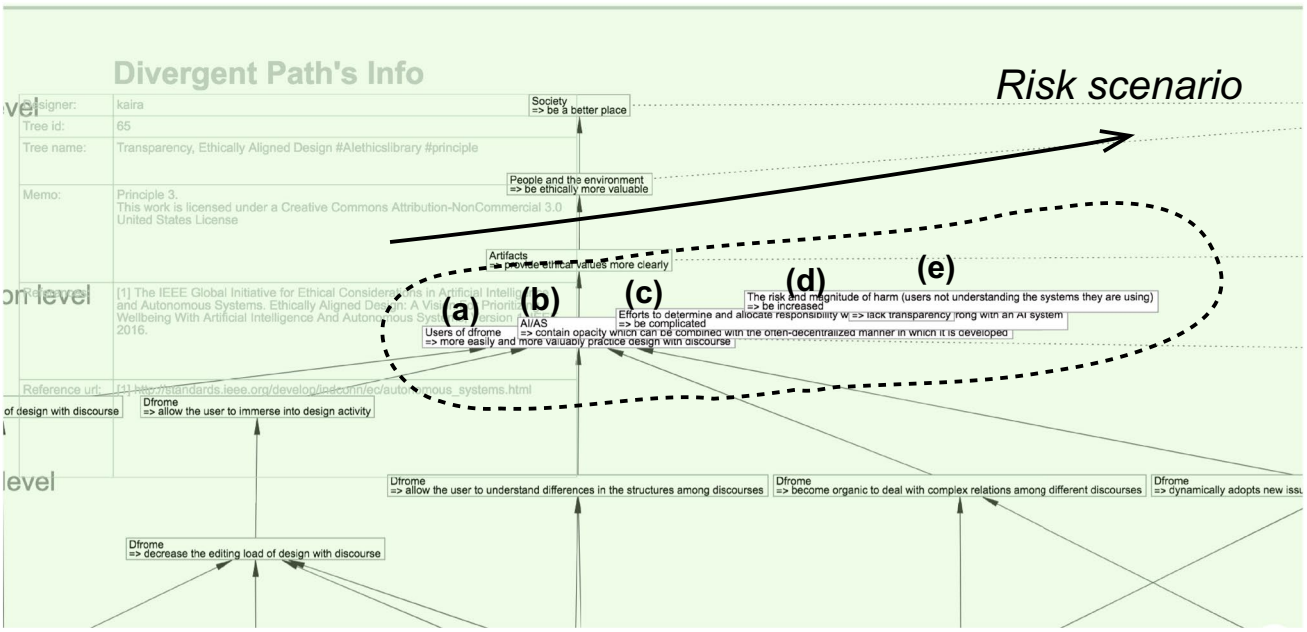

Fig. 21 A case of path recommendation for finding risk scenarios of dfrome are: (a) users of dfrome ⇨ more easily and more valuably practice design with discourse; (b) AI/AS ⇨ contain opacity which can be combined with the often decentralized manner in which it is developed; (c) Efforts to determine and allocate responsibility when something goes wrong with an AI system ⇨ be complicated; (d) AI/AS ⇨ lack transparency; (e) The risk and magnitude of harm (users not understanding the systems they are using) ⇨ be increased (IEEE 2016); rank is 1

understood that dfrome as a system (A) is composed of an interactive engine (B), a cloud environment accumulating various knowledge bases including AI ethics (C), and an investigation engine (D) (Fig. 20). The ethical scenario of dfrome is based on the idea that society would be a better place if ethically designed artifacts are more clearly realized and the same artifacts change their surroundings to be more ethically valuable, (Fig. 20 (E)).

Then, we often get asked if dfrome can build a neutral knowledge base of ethics. Our answer to this is that it is impossible and we also do not aim to achieve it. Rather, we aim to provide new awareness by clarifying the differences of understanding between users and encountering them.

However, applying a scenario recommendation of dfrome to find a risk scenario, another problem was suggested (Fig. 21). Here, we query a description of item (a) in Fig. 20 that is "Users of dfrome are changed to more easily and more valuably practice design with discourse". From the result, we were able to understand that dfrome might increase the magnitude of harm [Fig. 21(e)], because it can cause the lack of transparency when it enhances the co-creativity in a decentralized way.

Hence, dfrome were able to find future tasks for improving itself by applying the idea of dfrome. For example, with regards to transparency, we plan to implement a function that visualizes the history of edit logs including the recommendation results, so that the reference relation of the design idea can be clearly confirmed.

7 Discussion

Through the above-mentioned cases, we can confirm that dfrome can promote engineers' practice of ethical design.

First, by standardizing and visualizing differences in structures, the ethical design theory redefined the differences between the structures as the differences in how they appeared in our perspectives. This effect is useful for further ethical design because we can clearly understand whether the ethics level is expressed and where the gap exists. For example, the importance of considering wide influences for the outer environment can easily be achieved by seamlessly considering a number of upward branches. And, it is also useful to consider more creative solutions using the ideas that can complement the designer's own ideas by spreading the branches.

Second, regarding the tree distances, we could confirm that users can investigate whether there exist any similar ideas in the knowledge base. Furthermore, we can expect to complement the viewpoint by performing an interpretation of why some idea was calculated at a distance. For example, we can consider why paragraph vectors are distant from the responsibility, and we may find that one reason is the probability that they can produce large amounts of materials automatically without human controls, which can be a trigger for further ethical designs.

Finally, we can say that the path recommendation dynamically and clearly connects AI ethics to AI technology. Furthermore, because of the application of the ethical design theory that requires a description of the chain of changes, it became possible to confirm the consequences and spread the path that could not be easily imagined. For example, we confirmed that paragraph vectors are relevant to human rights, and dynamic mode decomposition is relevant to privacy protection. We are also able to investigate whether honoring human rights is what is most desirable. Additionally, dfrome itself can evolve by self-applying its functions and finding future tasks.

Furthermore, through these practices, it is possible to generate engineering-based feedback for AI ethics, and AI ethics will be reconstructed to be a more practical one rather than an empty theory. For example, AI ethics studies in the future can consider whether the AI ethics under investigation can be complementary at higher levels in the hierarchical representation of artifacts. The example of discussion with humanities, social sciences, and fine arts experts showed that dfrome is useful for bidirectional collaboration and such collaboration is necessary. Furthermore, this discussion example demonstrated that our study is open to meta-ethical discussion and how to practice it.

Additionally, we discussed with ethicists at 32nd Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2018) (Sekiguchi and Hori 2018a). We received feedback that our study is insightful from the viewpoint of ethicists. Furthermore, we discussed with social studies experts at Society for Social Studies of Science Annual Conference (4S Sydney) (Sekiguchi and Hori 2018b) about the difference between dfrome and the science interpreter. We received feedback that our study is significant even for them.

8 Conclusion

In this research, we implemented a tool for use with knowledge base of AI ethics named dfrome; this will help promote the practice of ethical design theory by AI engineers. We also applied a knowledge liquidization and crystallization model for building dfrome as a dynamic tool.

Our tool can make ethical AI design clearer and more standard. It can also make the process more ethical by supporting consideration of relationships among AI ethics and design ideas and by recommending scenario paths connected to the ethics level.

The three cases introduced in this paper made it certain that dfrome can deal with AI ethics both synthetically and semantically, and seamlessly connect AI ethics to AI technology. Therefore, we conclude that it can support AI engineers to more clearly and more easily investigate and design from the ethics level.

In the future, we plan to further develop dfrome, for example, by increasing transparency of processing and editing as discussed in Sect. 6.3.5, automating the tree description, and dealing with the data as a time series to consider the feedback of changes in trees. We also plan to

generate and collect concrete cases, and keep on collaborating among the experts in different sectors including humanities, social sciences, and fine arts experts to realize mutual improvement.

Acknowledgements The authors wish to acknowledge Dr. Hiroshi Nakagawa and Dr. Mayu Terada for their help in discussing drone delivery, drone highway and drone swarm. This work was partially supported by JSPS KAKENHI Grant Number JP18K18434.

References

Afacan Y, Demirkan H (2011) An ontology-based universal design knowledge support system. Knowl Based Syst 24(4):530–541. https://doi.org/10.1016/j.knosys.2011.01.002

AI Network (2017a) The conference toward AI network society, draft AI R&D guidelines for international discussions. http://www.soumu.go.jp/main_sosik/joho_tsusin/eng/Releases/Telecommunications/170728_05.html (Tentative Translation)

AI Network (2017b) The conference toward AI network society, report 2017: Toward promotion of international discussions on AI networking: overview. http://www.soumu.go.jp/main_sosik/joho_tsusin/eng/Releases/Telecommunications/170728_05.html

AI Network (2017c) The conference toward AI network society, socioeconomic impact of AI networking: Preliminary assessment. http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000067.html (Attached Paper 3 of Report 2017: Toward Promotion of International Discussions on AI Networking) **(in Japanese)**

Chapman C, Pinfold M (1999) Design engineerina need to rethink the solution using knowledge based engineering. Knowl Based Syst 12(5):257–267. https://doi.org/10.1016/S0950-7051(99)00013-1

Finke RA, Ward TB, Smith SM (1992) Creative cognition. MIT Press, Massachusetts

FLT (2017) Future of life institute, Asilomar AI principles. https://futureoflife.org/ai-principles/

Freedman B, Kahn Jr PH, Boring A (2006) Value sensitive design and information systems. In: Zhang P, Galletta DF (ed) Human–computer interaction and management information systems: foundations. M. E. Sharpe, pp 349–372

Guu K, Miller J, Liang P (2015) Traversing knowledge graphs in vector space. In: Proceedings of the 2015 conference on empirical methods in natural language processing, Lisbon, pp 318–327

Hori K (2004) Do knowledge assets really exist in the world and can we access such knowledge?; knowledge evolves through a cycle of knowledge liquidization and crystallization. In: Grieser G, Tanaka Y (eds) Intuitive human interfaces for organizing and accessing intellectual assets, Lecture Notes in Artificial Intelligence. Springer, pp 1–13

IEEE (2016) The IEEE global institute for ethical consideration in artificial and autonomous systems, Ethically aligned design: a vision for prioritizing wellbeing with artificial intelligence and autonomous systems, version 1. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

IEEE (2017) The IEEE global institute for ethical consideration in artificial and autonomous systems. Ethically aligned design: A vision for prioritizing wellbeing with artificial intelligence and autonomous systems, version 2. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Japan (1946) The constitution of Japan. http://www.japaneselawtranslation.go.jp/law/detail_main?id174

Jassowski M, Thirunahari AS (2018) Drone swarm for increased cargo capacity. In: United States Patent Application, 20180188724

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in neural information processing systems, vol 25. Curran Associates, Inc., pp 1097–1105

Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on international conference on machine learning, vol 32, JMLR.org, ICML'14, pp II-1188–II-1196

Lee J, Han S (2010) Knowledge-based configuration design of a train bogie. J Mech Sci Technol 24(12):2503–2510. https://doi.org/10.1007/s12206-010-1002-3

Miller J, Friedman B, Jancke G, Gill B (2007) Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. In: GROUP '07 proceedings of the 2007 international ACM conference on supporting group work, Sanibel Island, pp 281–290

Ministry of Land, Infrastructure, Transport and Tourism (2015) Japans safety rules on Unmanned Aircraft (UA)/Drone. http://www.mlit.go.jp/en/koku/uas.html

Myung S, Han S (2001) Knowledge-based parametric design of mechanical products based on configuration design method. Expert Syst Appl 21(2):99–107. https://doi.org/10.1016/S0957-4174(01)00030-6

Nakakoji K (2007) From interfaces to interactions: an overview of SIGHI (1001 SIG Nights). IPSJ Magazine, pp 202–203 **(in Japanese)**

Neuron Data Inc (1996) Neuron data elements environment v2.1—getting started. http://www.imn.htwk-leipzig.de/~sunpool/software/ee21/getstart/getstart.pdf

Nickel M, Murphy K, Tresp V, Gabrilovich E (2016) A review of relational machine learning for knowledge graphs. In: Proceedings of the IEEE, pp 11–33

Oka T, Morimoto M (2015) An extraction and recognition method for partially hidden objects. In: 2015 international conference on informatics, electronics vision (ICIEV), pp 1–4. https://doi.org/10.1109/ICIEV.2015.7334055

Oka T, Morimoto M (2016) A recognition method for partially overlapped objects. In: 2016 world automation congress (WAC), pp 1–4. https://doi.org/10.1109/WAC.2016.7583005

Řehůřek R, Sojka P (2010) Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, ELRA, Valletta, Malta, pp 45–50. http://is.muni.cz/publication/884893/en

Schmid PJ (2010) Dynamic mode decomposition of numerical and experimental data. J Fluid Mech 656(August):5–28. https://doi.org/10.1017/s0022112010001217

Sekiguchi K (2017) dfrome; website for design from the ethics level. https://www.dfrome.com

Sekiguchi K, Hori K (2018a) Can AI involve AI engineers in ethical design activities? In: Proceedings of the 32nd annual conference of the Japanese Society for Artificial Intelligence (JSAI2018), Kagoshima **(in Japanese)**

Sekiguchi K, Hori K (2018b) Organic and dynamic library of AI ethics for engineers. In: Proceedings of the society for social studies of science annual conference (4S Sydney), Sydney

Sekiguchi K, Tanaka K, Hori K (2009) "design with discourse" to design from the "ethics level". In: Družovec TW, Jaakkola H, Kiyoki Y, Tokuda T, Yoshida N (eds) Volume 206: information

modelling and knowledge bases XXI, Frontiers in Artificial Intelligence and Applications. IOS Press, pp 307–314
Simon HA (1996) The sciences of the artificial, 3rd edn. MIT Press, Massachusetts
Spiekermann S (2016) Ethical IT innovation: a value-based system design approach. CRC Press, Boca Raton, p 220
Terada M (2018) Legal study on drone highway. Inf Netw Law Rev 16:31–49 (**in Japanese**)
The Ethics Committee, The Japanese Society for Artificial Intelligence (2017) Open discussion. http://ai-elsi.org/archives/628
Tomiyama T (2016) Function allocation theory for creative design. Procedia CIRP 50:210–215. https://doi.org/10.1016/j.procir.2016.04.060, 26th CIRP Design Conference
Wang K, Nickerson JV (2017) A literature review on individual creativity support systems. Comput Hum Behav 74:139–151. https://doi.org/10.1016/j.chb.2017.04.035
Wittgenstein L (1969) On certainty, Basil Blackwell, pp 44–44e. First Harper Torchbook edition published 1972
Yanmaz E, Yahyanejad S, Rinner B, Hellwagner H, Bettstetter C (2017) Drone networks: communications, coordination, and sensing. Ad Hoc Netw 68:1–15
Yoshikawa H (1979) Introduction to general design theory. J Jpn Soc Precis Eng 45(536):906–912. https://doi.org/10.2493/jjspe1933.45.906 (**in Japanese**)
Yoshikawa H (1981) General theory of design process. J Jpn Soc Precis Eng 47:46–51 (**in Japanese**)
Yoshikawa H (2008) Introduction to theory of service engineering: framework for theoretical study of service engineering. Synthesiology 1:111–122 (**in Japanese**)