Capturing stakeholders values and preferences regarding algorithmic systems

Marceau Nahon, Aurélien Tabard, Audrey Serna

Capturing stakeholders values and preferences regarding algorithmic systems
Capturer les valeurs et préférences des utilisateurs face à un système intelligent

MARCEAU NAHON, University of Lyon, UCBL, INSA Lyon, CNRS, LIRIS, UMR5205 and Sorbonne Université, France

AURÉLIEN TABARD, University of Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR5205, France and Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, France

AUDREY SERNA, University of Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, France

Complex-decisions made by algorithmic systems should embed interests and values of the different stakeholders involved in the making of these systems. In this paper, we investigate how users' preferences can be captured in order to incorporate more ethical considerations in the design of such systems. Adopting a value sensitive design approach, we propose a method to measure how much the context can impact the differences between the declared importance of values and decisions made in a given situation. We developed a survey tool that enable to measure importance of values in either absolute or relative ways, comparing pairs of values in specific situations. We conducted a preliminary study to test our survey tool with fifteen participants in the context of Smart Grids. Our results show differences in the way participants estimate values and underline the interest of capturing users preferences in context.

Les décisions complexes prises par les systèmes algorithmiques devraient intégrer les intérêts et les valeurs des différentes parties prenantes impliquées dans la création de ces systèmes. Dans cet article, nous étudions comment les préférences des utilisateurs peuvent être mesurées afin d'intégrer des considérations plus éthiques dans la conception de ces systèmes. En adoptant une approche de conception sensible aux valeurs, nous proposons une méthode pour mesurer l'impact du contexte sur les différences entre les préférences déclarées et les décisions prises par les utilisateurs dans une situation donnée. Nous avons développé un outil d'enquête qui permet de mesurer l'importance des valeurs de manière absolue ou relative, en comparant des paires de valeurs dans des situations spécifiques. Nous avons mené une étude préliminaire pour tester notre outil d'enquête auprès de 15 participants dans le contexte des Smart Grids. Nos résultats montrent des différences dans la façon dont les participants estiment les valeurs et soulignent l'intérêt de capturer les préférences des utilisateurs en contexte.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**.

Additional Key Words and Phrases: Value Sensitive Design, Ethics, Artificial Intelligence, Smart-grids, Survey tool

Mots Clés et Phrases Supplémentaires: Design sensible aux valeurs, Ethique, Intelligence Artificielle, Smart-grids, Outil d'enquête

Reference:
Marceau Nahon, Aurélien Tabard, and Audrey Serna. 2024. Capturing stakeholders values and preferences regarding algorithmic systems. *IHM '24 : Actes étendus de la 35^{e} conférence Francophone sur l'Interaction Humain-Machine, March 25–29, 2024, Paris, France*

1 INTRODUCTION

Systems based on algorithmic treatment of data have been spreading alongside the development of Information and Communication Technology (ICT). Often grouped under the umbrella of Artificial Intelligence (AI) for their ability to make complex decisions, these algorithmic systems are generally programmed for specific goals. These goals and the technical implementation embed the values and interests of the organizations funding and developing it, of designers,

developers, and other stakeholders involved in the making of these systems. We are interested in understanding value sensitive design approaches could help incorporate more ethical considerations in the design of such systems [13, 20].

The works on Ethics in AI distinguish several approaches related to the integration of ethics consideration into these systems. *Ethics by Design* refers to the technical integration of ethical reasoning capabilities as part of the behavior of the system, whereas *Ethics in Design* refers to the engineering methods that support the analysis and evaluation of the ethical implications of AI systems [10]. In this paper, we adopt the *by-design* approach and we explore how to surface the values and general preferences of stakeholders on how algorithmic systems should behave. We focus on so-called "smart grids" scenarios, in which energy production and consumption must be aligned. We leverage a multi-agent system that seeks to model the variety of actors and reach decisions that optimizes individual preferences under global constraints. This optimization relies on arbitrage decisions made by the system and the people driving it.

The arbitrage decisions made by the algorithmic systems we are interested in can happen during the design phase of the system, while it runs, or even after some decisions when correcting unwanted behaviors, covering the full life cycle of AI technologies. As underlined in value sensitive design approaches, these decisions are context-dependent since different contextual variables impact the way values are understood by people. They are based on preferences, assumptions, values of individuals and groups, and sometime depend on external constraints from the state, companies, or local contextual elements such as weather, a sick person at home, living conditions, etc.

Inspired by value sensitive design, we present a survey approach to capture stakeholders preferences when challenging decisions about system behavior must be made. In particular, we are interested in understanding how much system decisions are in line with stakeholders preferences. We also want to explore how much the context can impact the decisions made and the differences between the declared importance of values and decisions in a given situation. We conducted a preliminary study to test our survey tool with fifteen participants.

2 RELATED WORK

2.1 Value sensitive design

The study of user values in AI has been explored in various domains [1, 7, 14]. This scholarship either adopts a very general point of view (e.g., questions about trust in AI [7]) or focuses specific application domains (e.g., automatic tax fraud identification systems [14]). Recent work started investigating the application of Value Sensitive Design (VSD) for AI technologies [19, 20]. They underline two challenges inherent to AI systems: the difficulty of understanding how these systems work for humans and the ability of these systems "to adapt themselves in ways that disembody the values embedded in them by VSD designers" [20].

2.2 By-design values in a smart grid context

We investigated more systematically research on values that would be relevant in the context of energy consumption and smart-grids. We focused on 12 relevant articles. Two were selected because they are references in Value Sensitive Design [7, 11]. We also selected specific cases of Value Sensitive Design in order to broaden the corpus of values. Five articles were selected because studying values in preferences in the context of AI [1, 2, 14, 18, 21]. Finally, two articles cover the smart grids application domain [5, 9].

Since we are interested in identifying values that can shape the behavior of an intelligent system, we decided to study a subset of values that the system can handle to make decisions that we call **by-design values**, referring to the *ethics by-design* approach. We believe that these **by-design values**, such as inclusiveness or affordability, represent

interesting arbitrage decisions or trade-offs to make. On the contrary, *in-design values* such as accountability or safety, represent values that can be evaluated but are not relevant for arbitrage decisions.

After a categorization inspired by Swchartz's [17] and a final selection of relevant by-design values we went from 75 to 6 values. We detail the values selected and our method below.

3 METHOD

We proceeded in three steps to create a tool to capture user preferences in a smart-grid context: i) selection of values, ii) creation of situations that confront each pair of values, iii) implementation of survey tools.

3.1 Selection of values

Literature review. At first, we wanted to gathered values relevant to the smart grids context as described above, with twelve articles. From these articles, we gathered a total of 75 values. After a semantic clustering, we went from 75 to 34 values. After that, we kept only the values that we considered relevant for our application domain, dropping to 25 values.

Categorization of values. We categorized the remaining values drawing inspiration from Swchartz's categorization of values [17]. Table 1 presents a simplified version of our categorization. For each value, we indicate the source articles.

Final selection. Since we focus on **by-design values** (on which the system can make trade-offs shape), we decided to put aside *in-design values* (non-negotiable ones). In the end, we established a list of nine *by*-design values which are :

- From **Self-determination**: Autonomy, Self-knowledge
- From **Stakeholders welfare**: Security of Supply, Affordability, Well-being, Privacy
- From **Equity**: Inclusiveness
- From **Global Welfare**: Environmental Sustainability

The **Self-determination** values (Autonomy and Self-knowledge) are difficult to put in confrontation with the other values of the list, as they are at another level. Indeed, we can intuit that **Stakeholders welfare**, **Equity** and **Global welfare** values are easy to confront as they quantify the alignment of the system regarding its results. On the other hand, the **Self-determination** values measure how the system interacts with the user and makes them intervene within the decision process, which is something different than what the system would suggest to the user. For what comes next, we decided not to keep these values. In the end, we kept the six following values and suggested ourselves a definition for each:

- **Inclusiveness:** The system promotes inclusion and equity for all users
- **Environmental sustainability:** The system minimizes the environmental impact
- **Privacy:** The system preserves users' privacy and personal information
- **Security of supply:** The system meets the user's primary and necessary energy requirements
- **Well-being:** The system responds to energy demands that exceed primary needs and are related to comfort
- **Affordability:** The system provides financially accessible services to users

3.2 Creation of situations

Situation. Drawing inspiration from the Value Sensitive Method called "Value Scenario" [12], we created situations that confront pairs of values to capture user preferences (15 situations in total, one for each pair of values). The idea is

Value	Category
Productivity [3] Efficiency [9] Reliability [9, 21] Machine Autonomy [14] Utility [6]	Machine efficiency
Privacy [1–3, 6, 8, 9, 14, 21] Security [3, 14] Safety [9] Accountability [18, 21]	Protectability (Schwartz Security)
Explainability [14] Transparency [1, 2, 8, 21] Usability [6] Disclosure [13]	Intelligibility
Autonomy [3, 6, 21] Self-knowledge [3]	Self-determination (Schwartz Self-direction)
Security of supply [5] Affordability [5] Well-being [21] Privacy [6]	Stakeholders welfare (Schwartz Benevolence)
Diversity, non-discrimination and fairness [1, 2, 4, 6, 9, 14, 21] Inclusiveness [5]	Equity (Schwartz Benevolence-Universalism)
Respect for law and public interest [2] Environmental sustainability [5, 9, 21] Trust [21]	Global welfare (Schwartz Universalism)

Table 1. The 25 final values and their respective category relevant for the smart grids context. Each value has been extracted from the literature analysis.

two present to survey participants, a scenario confronting two values and ask them to chose among the two suggested actions (each representing one of the two values). We relied on the pairwise comparison method [15]. A situation includes a statement and two options. The statement provides context and describe a case in which two values conflict. The two options represents the actions that can be chosen, each representing one of the two values. Here is an example with our Inclusiveness/Affordability situation:

- **Statement**: During a winter weekend, you have invited some friends over to watch TV/play video games/other similar activity, when the system informs you that some of your neighbors have not been able to heat as much as they wanted to, and that their comfort is significantly lower than most of the other residents. You have the option of buying energy (and therefore paying more), which will enable your neighbors to consume energy from the local grid to boost their comfort.
- **First Option - Inclusiveness**: Buy energy in order to watch TV/play video games/other similar activity.
- **Second Option - Affordability**: Watch TV/play video games/other similar activity as planned.

Limitation and biases. The first difficulty was to get close to "real reactions". The situations must be at the same time precise enough for a survey participant to relate to, and general enough for every participant to relate to. Another problem is the personal sensitivity of individuals. This was a within-participants limitation we had to take into account. While theoretically reducing heating or air conditioning can result in similar energy consumption for comfort, individual sensitivity to heat or cold can skew the results. For instance, if we suggest lowering the heating, those less sensitive to

cold could distort the findings. Another point is that the formulation of the statement can include bias. In the statement, we want the survey participants to be well aware of the consequences of both actions. However, we must not make them feel guilty if they choose one more "individual" value over a more "collective" one. Another bias was the status quo bias: a preference for the current situation [22].

3.3 Tool presentation

We designed an adaptable and configurable survey tool to capture user preferences. The tool can be adapted to various survey strategies, which enables researchers to carry out different surveys. The tool is composed of three modules and a menu used to configure (parameters and values studied) and launch the survey (Figure 3).

The first module (*situation* module) presents the situations defined previously to survey participants and ask them to choose an option, as illustrated in Figure 1. The researcher can customize the survey with parameters that determine the information presented to the user (it is possible to choose if the values beneath each option are displayed or hidden, and the difficulty and relevance are asked or not). When choosing an option instead of another, we consider that it means the participant prefers the corresponding value instead of the other one, in this given situation. In function of the difficulty of decision chosen by the participant, we can weight their preference.

The second module (*slider* module) allows to capture the importance of values by asking participants directly what importance they give to each chosen value (Figure 2). The six values are displayed with their definition. Each value has a slider that indicates its importance. The sliders offers seven values from "very little importance" (*très peu importante*) to "very important" (*très importante*).

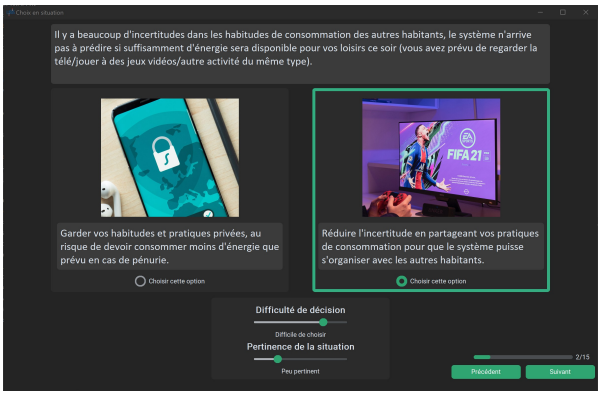

Fig. 1. ***Situation* module**: in this case the difficulty of decision (*difficulté de décision*) and the relevance of the situation (*pertinence*) are both asked to the user, the values stay hidden. In this case, the user chooses the second option, they consider that it is difficult to choose and that the situation is not really relevant.

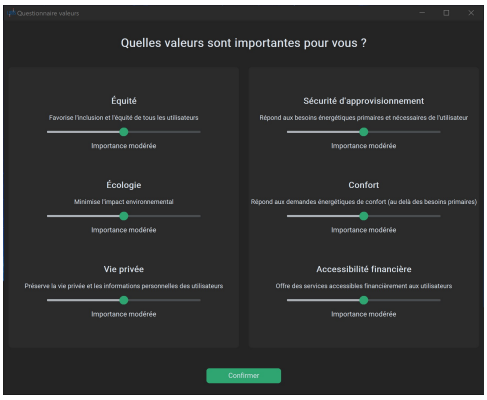

Fig. 2. ***Slider* module**: the six selected values are displayed and the user has to indicate their importance. What was asked during the preliminary study is how much each value is important for them in the case of smart grids.

Finally the third module (*agreement* module) is a variant of the first one: it presents a situation and explains that the system made a choice and is about to act. The participant is then asked to which degree they agree with this decision.

4 PRELIMINARY STUDY

4.1 Study design

In order test our survey tool, we conducted a small scale survey of fifteen participants. We wanted to study users' preferences in terms of values regarding smart grids. For this purpose, we added a menu to configure the survey (Figure 3) and a questionnaire to the survey tool to gather data about the participant (Figure 4). The idea behind the questionnaire is to have an idea on potential correspondences between some profiles and some value preferences. In our case, we defined questions relative to the energy consumption, as we considered it relevant in the case of smart grids.

We address three research questions in this study:

- In regard to the survey tool : does the tool can help designers in capturing users preferences in terms of values?
- Regarding the context of the smart grids: Is the importance of values context-dependent? what is the relative importance of these values?
- Regarding the broader context of users preferences toward intelligent systems: Is there an alignment between declarative judgment regarding "generic" values (that are not instantiated in a specific context) and actions in specific situations?

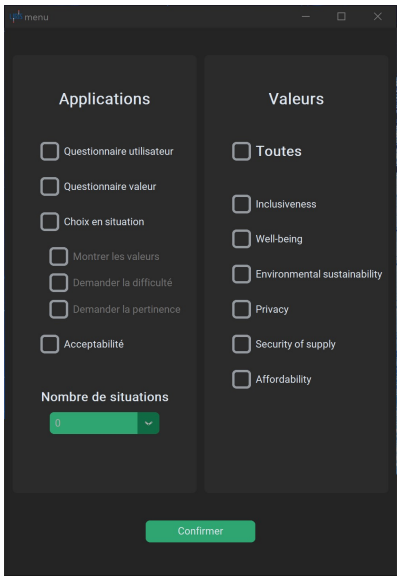

Fig. 3. **Menu**: configuration interface, selection and parameters on the left part, values selection on the right part.

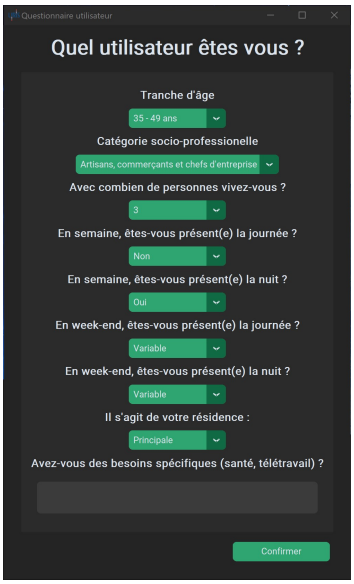

Fig. 4. **Questionnaire**: questions about the participant profile with regard to smart grids.

4.2 Procedure

The fifteen participants were students between 22 and 26 years old, most of them (13/15) in a field close to algorithmic systems (robotics, AI, or engineering degrees), and one close to the field of values (philosophy degree). The study was structured as a structured interview for in which the main author guided the participants through a survey. The interviews took place via video-conference. The main author displayed the survey tool on his screen. In order to give as

much freedom to the participants, they were also given the control of the screen. Interviews were conducted as follows: i) a brief explanation of the interview, ii) launch of the survey, iii) filling the demographic questionnaire, iv) answering questions about overall preferences, v) answering questions about specific situations (Figure 5). In order to not to take too much time, we decided to present only one situation for each pair of values. This captured an idea of participants preference but with a lower precision compared to a survey with several situations for each pair.

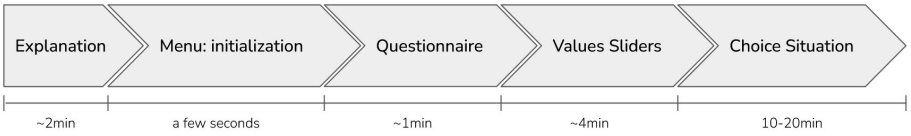

Fig. 5. Survey proceeding

4.3 Data collection

The data obtained by each interview was collected in real time on a csv file. The file gathered:

- From the *Slider module*: For each situation, a score from 0 to 1 for each value depending on the responses. The sliders have 7 numerical values from 0 to 1.
- From the *Situation module*: the id of the situation, the chosen value (0 or 1) and a score from 0 to 1 for the difficulty and relevance (5 numerical values). The score for the two involved values goes from 1-0 to 0.5-0.5 depending on the difficulty.

4.4 Data analysis

In order to analyze the data, we compare participants' preferences obtained by the *Slider module* and the *Situation module*. To do so, we relied on pairwise comparison [15] analysis method. In this approach, each participant has an individual matrix of comparison that we can use to calculate the relative importance of each value. We built a matrix of comparison for each module. We used the principal eigenvector of each matrix to attribute a final score for each value [16].

5 RESULTS

For each participant, we first analyzed the relative importance of values in comparison to the others thanks to the results of the situation module using the pairwise comparison approach [15]. Figure 6 presents the results of one the participants. This participant estimated well the importance of every value except two: the participant very clearly underestimated the importance of the inclusiveness and overestimated the importance of the security of supply.

Then we analyzed results from a global perspective thanks to the average profile obtained by computing the mean individual matrices, as illustrated in Figure 7.

At first, we note that this average profile is balanced and that the errors are in general rather low. We can see that participants initial estimates for some values are not aligned to their later responses in the survey. For instance, Security of supply loses three places (overestimated in the absolute) and well-being wins one (underestimated). In addition, we can see that the absolute difference for Inclusiveness, Environmental sustainability, Privacy and Affordability is very low and the mean difference for these values is not very high. This means that the participants manage to properly estimate the importance of these values, and that there is more or less as much overestimation as underestimation.

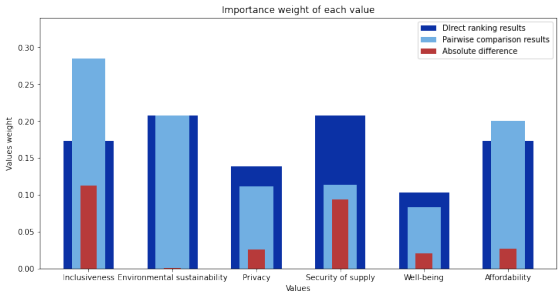

Fig. 6. **Results of a participant**: in dark blue the results of the direct ranking, in light blue the results of the main application, in red the difference between both.

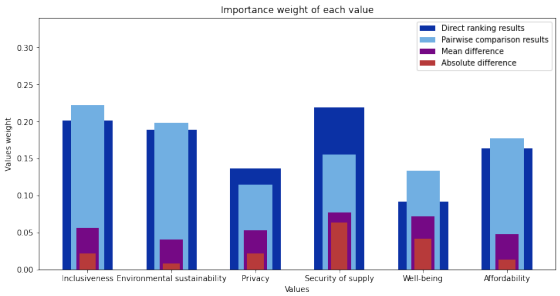

Fig. 7. **Global results**: in violet the mean difference of individual results between the main application and the direct ranking. In red the absolute difference: we take into account the fact that some users overestimate a value while others underestimate it, which compensates.

6 DISCUSSION AND CONCLUSION

To conclude, we can assume that our selection of values seems to be appropriate. Indeed, we saw with the global results that for our average profile no value was left aside, and they were all important. We can also assume that the applications and our pairwise comparison works well when we see that the difference between the results of the pairwise comparison and the direct ranking is rather low. However, the issue concerning Security of supply and Well-being remains. We can assume that when having to rank values without context, participants tend to put i) security of supply in first position because it seems to be the most important thing for a smart grid, ii) well-being in last position because it seems to be the least important, and the fact that this value is more "individual" probably contributes. On the other hand, when the participants are facing a situation, they tend i) to make concessions on the Security of supply for the benefit of the Inclusiveness or the affordability, ii) not to completely leave aside their Well-being. Moreover, the great difference may also be due to the fact that the arbitrary decision of what belongs to primarily needs or comfort is no necessarily shared by participants. An unexplored lead to deal with it consisted in an additional question in the questionnaire. The participant would be asked to rank several (3-5) appliances by importance. Then, the most important appliance would be used in the Security of supply situations, and the least important would be used in the Well-being situations. This way, we would ensure that the distinction is effective for the participant. We conducted the conceptual investigation and started the empirical investigation with the preliminary study. The next step would be to conduct a larger scale study to complete the empirical investigation. After, that, only the technical investigation would remain, it would consist in the implementation of the values in the system.

ACKNOWLEDGEMENTS

We warmly thank the participants of our study, as well as Rémi Chaput and Matthieu Guillermin for their feedback. This work was particially funded by ANR (Agence National de la Recherche) AAPG 2022, grant ANR-22-CE23-0028-01. We also thank the LABEX ASLAN (ANR–10–LABX–0081), Université de Lyon, within the framework of the French program "Investissements d'Avenir" managed by the ANR.

REFERENCES

[1] Banu Aysolmaz, Rudolf Müller, and Darian Meacham. 2023. The public perceptions of algorithmic decision-making systems: Results from a large-scale survey. *Telematics and Informatics* 79 (02 2023), 101954. https://doi.org/10.1016/j.tele.2023.101954

[2] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The Values Encoded in Machine Learning Research. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 173–184. https://doi.org/10.1145/3531146.3533083

[3] Karen Boyd, Adam Rule, Aurélien Tabard, and Jim Hollan. 2016. Sharing, Human Values, and Computer Activity Tracking. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (San Francisco, California, USA) *(CSCW '16 Companion)*. Association for Computing Machinery, New York, NY, USA, 233–236. https://doi.org/10.1145/2818052.2869119

[4] Jenna Burrell, Zoe Kahn, Anne Jonas, and Daniel Griffin. 2019. When Users Control the Algorithms: Values Expressed in Practices on Twitter. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 138 (nov 2019), 20 pages. https://doi.org/10.1145/3359240

[5] Rémy Chaput. 2022. *Learning behaviours aligned with moral values in a multi-agent system: guiding reinforcement learning with symbolic judgments.* PhD Thesis. Université Claude Bernard Lyon 1 (UCBL). https://theses.hal.science/tel-04011033

[6] Bodong Chen and Haiyi Zhu. 2019. Towards Value-Sensitive Learning Analytics Design. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge* (Tempe, AZ, USA) *(LAK19)*. Association for Computing Machinery, New York, NY, USA, 343–352. https://doi.org/10.1145/3303772.3303798

[7] Hyesun Choung, Prabu David, and Arun Ross. 2022. Trust and ethics in AI. *AI SOCIETY* 38 (05 2022), 1–13. https://doi.org/10.1007/s00146-022-01473-4

[8] European Commission. 2019. *Ethics Guidelines for Trustworthy AI.* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[9] Tristan De Wildt, Emile Chappin, Geerten Van de Kaa, Paulien Herder, and Ibo Poel. 2019. Conflicting values in the smart electricity grid a comprehensive overview. *Renewable and Sustainable Energy Reviews* 111 (05 2019), 184–196. https://doi.org/10.1016/j.rser.2019.05.005

[10] Virginia Dignum. 2018. Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology* 20 (02 2018). https://doi.org/10.1007/s10676-018-9450-z

[11] Batya Friedman and David G. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination.* The MIT Press. https://doi.org/10.7551/mitpress/7585.001.0001

[12] Batya Friedman, David G. Hendry, and Alan Borning. 2017. A Survey of Value Sensitive Design Methods. *Foundations and Trends® in Human–Computer Interaction* 11, 2 (2017), 63–125. https://doi.org/10.1561/1100000015

[13] Batya Friedman, Peter Kahn, Alan Borning, Ping Zhang, and Dennis Galletta. 2006. *Value Sensitive Design and Information Systems.* https://doi.org/10.1007/978-94-007-7844-3_4

[14] Kimon Kieslich, Birte Keller, and Christopher Starke. 2022. Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data & Society* 9, 1 (2022), 20539517221092956. https://doi.org/10.1177/20539517221092956 arXiv:https://doi.org/10.1177/20539517221092956

[15] María Pérez-Ortiz and Rafal Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. (12 2017).

[16] Thomas Saaty. 2003. Decision-making with the AHP: Why is the principal eigenvector necessary. *European Journal of Operational Research* 145 (02 2003), 85–91. https://doi.org/10.1016/S0377-2217(02)00227-8

[17] Shalom H. Schwartz. 2012. An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture* 2 (2012).

[18] Zsofia Toth, Robert Caruana, Thorsten Gruber, and Claudia Loebbecke. 2022. The Dawn of the AI Robots: Towards a New Framework of AI Robot Accountability. *Journal of Business Ethics* 178 (07 2022). https://doi.org/10.1007/s10551-022-05050-z

[19] Steven Umbrello and Angelo Frank De Bellis. 2018. A Value-Sensitive Design Approach to Intelligent Agents. In *Artificial Intelligence Safety and Security*, Roman Yampolskiy (Ed.). New York, NY, USA: CRC Press, 395–410.

[20] Steven Umbrello and Ibo Poel. 2021. Mapping value sensitive design onto AI for social good principles. *AI and Ethics* 1 (08 2021), 3. https://doi.org/10.1007/s43681-021-00038-3

[21] I.R. van de Poel, T.E. de Wildt, E.T. Oosterlaken, and M.J. van den Hoven. 2022. *Ethical and societal challenges of the approaching technological storm.* European Parliamentary Research Service. https://doi.org/10.2861/61151

[22] Richard Zeckhauser and William Samuelson. 1988. Status Quo Bias in Decision-Making. *Journal of Risk and Uncertainty* 1 (02 1988), 7–59. https://doi.org/10.1007/BF00055564

A VALUES TABLE