What Is Ethical AI? – Design Guidelines and Principles in the Light of Different Regions, Countries, and Cultures

Sarah K. Lier
Leibniz University Hannover
lier@iwi.uni-hannover.de

Jana Gerlach
Leibniz University Hannover
gerlach@iwi.uni-hannover.de

Michael H. Breitner
Leibniz University Hannover
breitner@iwi.uni-hannover.de

Abstract

Artificial Intelligence (AI)'s impact on societies is positive and negative. Human well-being, self-actualization, human agency, and social cohesion come with challenges of overuse, underuse, and misuse of AI systems and social anxiety, ignorance, or erroneous data. An implementation of AI Ethics is expected to address these challenges. Literature includes general or specific guidelines for ethical AI, but country-, region-, and culture-specific categorizations are limited. We derive ethical AI key topics (KTs), design requirements (DRs), and design principles (DPs). We apply text mining and topic modeling analysis in a Design Science Research (DSR)-oriented approach. From 187 scientific publications, we deduce four KTs, 13 DRs, and 15 DPs. We identify four regions, countries, and cultures and apply cultural dimensions to assign a prioritization of the DPs. This ranking enables ethical AI realizations in different regions, countries, and cultures.

Keywords: Ethical AI, Cultural Dimensions, Design Principles, Design Science Research.

1. Introduction

AI system usage and AI research have increased. Scientists, professionals, companies, and others use AI systems for, e.g., predictions, autonomous decision-making, or decision support (Carter et al., 2020; Yu et al., 2018). Advanced AI systems are applied in various areas and often are assisted or supervised by humans (Bankins & Formosa, 2023; Dolganova, 2021). AI system usage impacts positive and negative societies (Paraman & Anamalah, 2023; Mikalef et al., 2022; Mirbabaie et al., 2022; Floridi & Cowls, 2019). Positive impacts are human agency, technological advancement for human well-being, self-actualization of individuals and groups, and societal cohesion. Negative impacts are overuse, underuse, or misuse of AI systems and induce fear, ignorance, misplaced concerns, and excessive societal reactions (Floridi & Cowls, 2019). Ethical AI ensures societal benefits and avoids misuse or underuse of these systems (Floridi et al., 2018). Incorporating ethical guidelines and principles into AI systems increases AI's fairness and responsibility (Paraman & Anamalah, 2023; Zhang et al., 2023; Arrieta et al., 2020). The benefits of ethical AI can be described by a society's usage, acceptance, and recognition of new opportunities (Paraman & Anamalah, 2023; Floridi & Cowls, 2019; Yu et al., 2018). Society's acceptance and adoption are prerequisites for AI systems (Paraman & Anamalah, 2023; Floridi & Cowls, 2019). Mikalef et al. (2022) address a potential loss of control of autonomies. Mirbabaie et al. (2022) address a conflict between AI and Ethics. This conflict comprises big data, AI autonomy, and protecting the rights of individuals and autonomies (Mirbabaie et al., 2022). Some scientists address ethical principles for AI systems (e.g., Bankins & Formosa, 2023; Nguyen et al., 2023; Paraman & Anamalah, 2023; Prem, 2023; Mikalef et al., 2022; Ryan & Stahl, 2021; Hagendorff, 2020; Peters et al., 2020; Floridi, 2019; Floridi & Cowls, 2019; Jobin et al., 2019; Yu et al., 2018). This research must be joined with guidelines and principles to address society's fear and upgrade the growing research (Bankins & Formosa, 2023; Seo & Thorson, 2023; Mirbabaie et al., 2022; Ryan & Stahl, 2021; Jobin et al., 2019; Yu et al., 2018).

We consider different regions, countries, and cultures and reinforce cultural relevance, diversity, and social inclusion. We use cultural dimensions for region, country, and culture selection according to Hofstede (2023; 2010) and consider the USA, Western Europe, China, and India as important for ethical AI investigation. We follow the DSR-oriented approach inspired by Vom Brocke et al. (2020) and Hevner & Chatterjee (2010). DSR is characterized by the flexibility to constantly change literature, a practice- and solution-oriented view, and innovation potential. DSR provides a practical and application-oriented approach to develop ethical AI systems that follow ethical principles and address the requirements and values of concerned stakeholders. By deriving DPs, the gap between ethical theory and practical implementation can

URI: https://hdl.handle.net/10125/107207
978-0-9981331-7-1

be closed (Hevner & Chatterjee, 2010). Developing our design guidelines and principles addresses efficiency, consistency, aesthetics, and ethical standards for AI systems (Hevner & Chatterjee, 2010). We use text mining and topic modeling to identify patterns and trends in literature, improve research results, and detect hidden information (Gerlach et al., 2022; Tong & Zhang, 2016). For our literature review, we follow Vom Brocke et al. (2015), Webster & Watson (2002), and Watson & Webster (2020). We develop a design artifact as design guidelines and principles for ethical AI. We derive ethical guidelines and principles from literature and assign them to the selected regions, countries, and cultures. We address the research questions (RQs):

RQ1: How ethical AI perspectives can be deduced with cultural dimensions for different regions, countries, and cultures?

RQ2: How design guidelines and principles can be developed, and how do they relate to different regions, countries, and cultures?

We consider the cultural dimensions and discuss the relevance of ethical AI. We follow a nine-step DSR. We identify KTs through topic modeling and text mining following a literature analysis. We deduce DRs through KTs and design guidelines and principles. We prioritize our DPs concerning regions, countries, and cultures. We adjust our results through three expert interviews. After an adaptation of our results and findings, we discuss implications and recommendations for theory and practice and present a further research agenda.

2. Theoretical Background

Ethical AI has no unified definition and depends on the definition of AI. Ryan & Stahl (2021) describe the property of an AI system to fully and correctly interpret datasets and learn and gain knowledge from data. Schrader & Ghosh (2018) consider AI as a complex system designed to train, learn, and think like humans. The ability of AI to emulate human decision-making can increase productivity but leads to security and ethical issues. This increases the attack surface for hackers by, e.g., encoding human biases and errors in AI systems (Berente et al., 2021). Ethical AI addresses challenges in privacy, bias, denial of autonomy, discrimination, transparency, uncertainty, and misuse of AI systems (Bankins & Formosa, 2023; Paraman & Anamalah, 2023; Yu et al., 2018). Ethical considerations must address society's fears of AI risks (Paraman & Anamalah, 2023; Yu et al., 2018). Ethics describes a philosophical discipline and science of human moral behavior that deals with values and norms (Schrader & Ghosh, 2018). Moral concerns concrete and factual behaviors, groups, or individuals. Ethics and morals establish ethical laws, foundations, and prohibitions in regions, countries, and cultures. Ethics and morals are independent of the AI description. However, they must be observed for the AI and its acceptance and success (AI, 2019). The description of ethical AI is using AI systems strictly obeying ethical principles and values (e.g., Schrader & Ghosh, 2018; Yu et al., 2018). The purpose is to ensure the development, implementation, and usage of AI systems that comply with ethical standards and positively impact individuals, societies, and the environment (Paraman & Anamalah, 2023; Mirbabaie et al., 2022; Schrader & Ghosh, 2018). Vallor (2016) highlights the relevance of ethical virtues in using technologies and develops aspects of the interplay of virtues and technologies, technological virtues, practices, environmental Ethics, and education. We refer to virtues and technologies and technological education. Various ethical considerations on AI can be identified in the literature (e.g., Mikalef et al., 2022; Ryan & Stahl, 2021; Hagendorff, 2020). Seo & Thorson (2022) note that ethical regulations are not static but flexible and need to be adapted or revised. The listed publications do not consider countries on which the results of principles for ethical AI are based. Some studies relate to normative Ethics due to responsibility, value-based orientation, avoidance of negative impacts, legal and political regulations, and social acceptance (e.g., Mirbabaie et al., 2022; Schrader & Ghosh, 2018; Yu et al., 2018). We follow normative Ethics because of the stated goals of ethical AI. Normative Ethics comprises different theories, as reflected in our country's selection (Schrader & Ghosh, 2018). We apply Hofstede's (2023; 2010) cultural dimensions to generate cultural differences in values, attitudes, and behaviors.

Table 1. Selected regions, countries, and cultures based on cultural dimensions, ethical theories, and population
*Average *[2]Cumulated

	Cultural dimensions							
Country	PDI	IDV	MAS	UAI	LTO	Essential ethical theories	Population	References
China	80	20	66	30	87	Confucianism	1430 Mio.	Roberts et al. (2022); Wu et al., (2020); Feldmann et al. (1999)
India	77	48	56	40	51	Hinduismus, Sikhismus Buddhismus, Jainismus,	1420 Mio.	Chatterjee & NS (2022); Marda (2018); Kalyanakrishnan et al. (2018)
USA	40	91	62	46	26	Utilitarism	340 Mio.	Joh (2022); Mancilla et al. (2022); Pesapane et al. (2018)
Western Europe	43*	65*	54*	69*	59*	Deontology	390*[2] Mio.	Roberts et al. (2022); Stahl et al. (2022); Pesapane et al. (2018)

Thus, we address RQ1. Hofstede (2023; 2010) defines five cultural dimensions. The power distance index (PDI) describes the extent of power relations in a culture (a high imbalance in the distribution of power means high power distance). The cultural dimension of individualism (IDV) describes the extent to which an individual's interests are subordinate (collectivism) or superior (individualism) in the group. Masculinity (MAS) describes the allocation of tasks within the culture. The uncertainty avoidance index (UAI) defines the handling of unknown dimensions, and long-term orientation (LTO) is directed toward short-term or long-term success (Hofstede, 2023; 2010). Our selection by cultural dimensions involves China, India, the USA, and Western Europe. Our decision was supported by the population size of each country and their technological progress in AI. We consider Western Europe as a region because these countries develop technological standards in cooperation. Standards on ethical AI within the countries relate to each other and are considered best practices. India and China have a high level of power. Power and authority are distributed from the top down in China and India. There is a division between power holders and society in China, so society's interest is not considered directly. Low hierarchies are pursued in Western Europe and the USA, and societal equality is strived for. Another difference is individualism in the USA and Western Europe, where freedom and personal responsibility are pursued, and collectivism in India and China. The four selected regions, countries, and cultures are leaders in (global) technology and AI development and have different legal frameworks. The laws and regulations related to Ethics and AI identify different approaches and procedures. Table 1 presents the regions, countries, and cultures with Hofstede's scores, ethical theories, approximate population size, and sample references. Western Europe is an average (rounded) of Austria, Belgium, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Portugal, Switzerland, Spain, and the United Kingdom (Ford & Jennings, 2020); see online Appendix A. The population size of regions, countries, and cultures is based on 2022. We cumulated the population size for Western Europe for each western country (Eurostat, 2023). The population size for China, India, and the USA is based on data from the United Nations (2022). Values for the cultural dimensions were calculated using Hofstede (2023). The ethical theories were not discussed.

3. Research Design

3.1. Design Science Research

To address RQ2, we apply the DSR framework based on Vom Brocke et al. (2020) and Hevner & Chatterjee (2010). We also follow the DSR scheme from Gregor et al. (2020). These DSR approaches focus on generating new knowledge in artifacts and solving real-world problems. The application and problem-oriented DSR approach can analyze changing issues and consider the state of literature and research. Optimizing the artifact and providing an understanding of the topic, the DSR approach creates an artifact to solve research challenges with accompanying analysis. DSR comprises iterative development by continuously adapting and improving the design artifact (Hevner & Chatterjee, 2010). It is possible to engage stakeholders and integrate theory and practice in developing and evaluating the artifact (Vom Brocke et al., 2020). We use text mining and topic modeling analysis to identify patterns and trends in literature and improve objective research results. A machine learning (ML) text mining tool extracts knowledge, reviews literature, and reveals hard-to-detect information (Tong & Zhang, 2016). Based on the clusters, we identify KTs, derive requirements from the KTs, and sort them by relevance to obtain DRs. We deduce DPs from the literature dataset and combine them with DRs. DRs are concrete requirements and specifications of the artifact's performances, properties, and functions (Gregor & Hevner, 2013). DPs provide to deal with DRs. This facilitates the clarity, transferability, and readability of DRs and DPs (Gerlach et al., 2022). We contribute our DSR artifact to level 2 (nascent design theory) according to the DSR description of Gregor & Hevner (2013). This implies

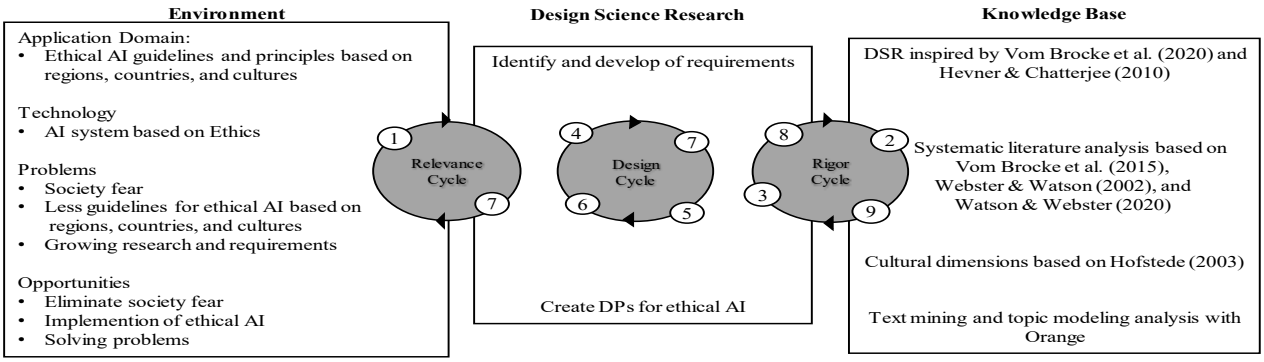

Figure 1. Research design inspired by Vom Brocke et al. (2020) and Hevner & Chatterjee (2010)

proposing more general artifacts, such as methods or DPs (Gregor & Hevner, 2013). Our artifact is formed by DPs considering user activities for ethical AI. DPs provide general guidelines to guide the design in a given direction and achieve quality (Gregor & Hevner, 2013).

The structure and progression of our research design are highlighted in Figure 1. Introduction and theoretical background define the problem formulation (step 1; relevance cycle). For step 2 (rigor cycle), we conduct a literature search and use a text mining tool to cluster and form KTs. In step 3 (rigor cycle), we identify initial requirements from our literature dataset. We derive DRs based on the initial requirements in step 4 (design cycle). We deduct DPs about our DRs in step 5 (design cycle). In step 6 (design cycle), we adjust the DRs and DPs by expert interviews. We assign the DPs to regions, countries, and cultures (step 7; relevance and design cycle). Our results will be discussed in step 8 (rigor cycle). We develop contributions for theory and practice and a further research agenda (step 9; rigor cycle).

3.2. Building a Knowledge Base

We follow the literature analysis inspired by Vom Brocke et al. (2015), Watson & Webster (2020), and Webster & Watson (2002). The literature review relates its advantages to summarizing the state of knowledge, identifying research gaps, and assessing the research quality (Vom Brocke et al., 2015). We perform a keyword-based search in SpringerLink, Web of Science, Elsevier, AIS Electronic Library, IEEE Explore, and ACM databases using the search string: "ethical AI" OR "ethical artificial intelligence" AND "guideline" OR "principle" OR "solution" during the period from January 2017 to May 2023. The keyword-based search identifies the timeliness of publications, emphasizes relevance and efficiency, and results in a comprehensive search. Disadvantages of this search include irrelevant results and limited coverage and contextualization (Watson & Webster, 2020; Vom Brocke et al., 2015). Therefore, we continue the literature search. After reviewing titles and abstracts, 193 papers were selected. We analyzed the full text and excluded 24 papers. Following Watson & Webster (2020) and Webster & Watson (2002), we added 23 papers in backward, forward, author, and Google Scholar similarity searches. The selection of literature was dependent on the added value of a paper (contribution), the quality of the journal or conference, i.e., white papers are not included, impact factors, argumentation and novelty of research results and findings, and citations, e.g., on Google Scholar and Scopus. Five papers were excluded after review. We included 187 papers in our final dataset. This dataset is used for our topic modeling and text mining analyses with the Python-based text mining tool Orange, proposed by Demšar et al. (2013). Our selection of the top-down method allows us to identify clusters from the dataset and interpret them as KTs. We adapted the text mining and topic modeling approach of Gerlach et al. (2022). First, we cleaned the data by deleting, e.g., titles, references. In the second step, the data are preprocessed by, e.g., removing punctuation marks, and then creating a keyword list with the most irrelevant words, e.g., the, and. Then, we created a word cloud to identify the most frequently used words. In step 4, we transformed the dataset by hierarchical clustering. Four clusters were identified. In step 5, we applied the topic modeling approach of Tong & Zhang (2016) to form four datasets that can be identified as KTs. The KTs have the advantage of revealing research gaps in the clusters (Gerlach et al., 2022; Tong & Zhang, 2016).

4. Results and Findings

4.1. Key Topics

KT1 focuses on human thinking, acting, and designing AI to be ethical (38 publications). In these publications, users conduct surveys, making the results user-oriented. This provides a view of the requirements of ethical AI from users and their needs. **KT2** includes developing AI systems with ethical considerations (42 publications). This implies a technological perspective and provides requirements from the developer's point of view. Cluster 3 (45 publications) describes guidelines for ethical AI from different countries. These principles are defined by the state or by authors who derived their principles from the perspective of a country. We use **KT3** to identify priorities of the selected regions, countries, and cultures. **KT4** deals with general requirements of ethical AI (40 publications). The graphic in the online Appendix B presents the top ten words from each KT on the vertical axis. The horizontal axis highlights the ratio of words, which varies across clusters based on the distributions of contributions.

4.2. Design Requirements

We analyzed the KTs and searched for existing requirements for the design of ethical AI. The requirements were sorted based on their relevance by, e.g., multiple nouns, and then included as a DR. **DR1** requires to support humans without replacing them. Ethical AI facilitates the execution of decisions and resultant experiences without replacing humans. Ethical AI needs the acceptance of humans in decision support (Bankins & Formosa, 2023; Ghotbi et al., 2022; Dolganova, 2021). **DR2** requires AI systems to be technically robust and secure. Ethical AI must work

based on user expectations and not be a security risk. Ethical systems must be able to protect data (Currie et al., 2020; Gerke et al., 2020). Data protection is a key requirement for AI systems (**DR3**). Sensitive data affects the personal data of individuals. Data must not be disclosed to third parties to ensure that trust in AI systems is not compromised. Ethical AI must handle data responsibly (Nguyen et al., 2023; Kaur et al., 2022; Gerke et al., 2020). **DR4** requires transparency in AI systems, decisions, and actions. Users must know how the system works and what expectations they can have. Transparency is important to identify damage caused by an attack and leads to trustworthiness and comprehension of ethical AI (Kaur et al., 2022; Green, 2018). Fairness, diversity, and non-discrimination are expected from **DR5**. Exclusion of an individual or group based on inborn or learned characteristics and factors that do not influence the decision must not occur (Paraman & Anamalah, 2023; Zhang et al., 2023; Dolganova, 2021; Currie et al., 2020). **DR6** requires assurance of economic and social well-being. Ethical AI must benefit people and society and add value. Integrating ethical AI into society is relevant for the next generations to have less fear and a responsible awareness of AI usage (Paraman & Anamalah, 2023; Chao, 2019; Green, 2018). **DR7** calls for awareness and encourages children and young people to use and build ethical AI. Their education can eliminate ethical doubt and promote attitudes toward AI (Ghotbi et al., 2022; Forsyth et al., 2021; Leimanis & Palkova, 2021). Including human emotions in models and data for ethical AI algorithms is necessary (**DR8**). Human-centered models facilitate ethical decision-making and the resolution of challenges in practice (Ho & Wang, 2021; Buenfil et al., 2019). **DR9** states that ethical AI should increase quality and be trustworthy. Quality refers to data that must be of high quality for analysis (Peters et al., 2020; Steimers & Bömer, 2021). **DR10** calls to filter disinformation that causes or seeks harm. Disinformation increases as technological advances; ethical AI can assist in its detection and prevention (Lange & Lechterman, 2021). **DR11** calls for avoiding data bias, e.g., caused by training datasets. Bias leads to exploitative, discriminative, and unethical decisions (Paraman & Anamalah, 2023; Naik et al., 2022; Carter et al., 2020; Mujtaba & Mahapatra, 2019). **DR12** calls for ethical AI not to harm anyone and to avoid harm in its development, deployment, and usage (Kaur et al., 2022; Leimanis & Palkova, 2021).

4.3. Design Principles

Based on the KTs and DRs, we derived DPs. **DP1** proposes AI systems based on human supervision and action. The user must be aware of the risks and limitations of an AI system, while an AI system must be designed to meet the users' requirements. Based on accelerated decision-making, existing staff can manage, learn how to use and handle AI systems (Bankins & Formosa, 2023; Dolganova, 2021; Carter et al., 2020). **DP2** is concerned with structured behavior during cyberattacks. An AI system needs to be resilient and recover from attacks that result in damage, remain fully functional, and not cause harm to a person. Results must be able to be reproduced (Kaur et al., 2022; Steimers & Bömer, 2021). In consideration of the General Data Protection Regulation (GDPR) (**DP3**) is important in data protection and building human trust in AI systems. This can be used to maintain data quality and integrity and thus protect data. Systems should integrate and act to GDPR (Kaur et al., 2022; Meske & Bunde, 2021). **DP4** describes the need for a sufficient comprehension of the performance and limitations of ethical AI. Users must know the information used to propose or reject a decision to avoid discrimination or bias. Training and engagement with AI systems are necessary (Paraman & Anamalah, 2023; Kaur et al., 2022; Nwafor, 2021). Access and sufficient availability in all parts of society are **DP5** solutions. Conditions of access and availability are important to establish equality. The consideration of datasets in data collection of society for inclusion and fair treatment is needed (Paraman & Anamalah, 2023; Chao, 2019; Green, 2018). It is important to provide children, young people, and adults with access to ethical AI, as **DP6** describes. For adults, the weakness is technological comprehension; for children and young people, the weakness is insufficient education on handling, using, and building AI systems. Courses or family members can educate adults. Children and young people can be informed through school. Informed students can evaluate hazards and risks and learn to handle, operate, and interact with ethical AI. Critical questioning is encouraged, and long-term value is added to sustain research in ethical AI (Ghotbi et al., 2022; Forsyth et al., 2021; Leimanis & Palkova, 2021; Wang et al., 2020). **DP7** describes the establishment of risk management. This can be reached by developing ethical AI by recognizing behaviors in a programmed manner and classifying them as hazards (Ghotbi et al., 2022; Rakowski et al., 2021; Wang et al., 2020). **DP8** considers the real-time assessment of data and models. This increases prediction probabilities or leads to faster detection of attacks. Real-time evaluation can be achieved by constantly retrieving and analyzing data. It is possible to allow employees or users to control data and models in time, viewing a system's live current states (McGregor et al., 2021). As described by **DP9**, optimization considers its advantage in avoiding data bias. Optimization of datasets reduces and eliminates discriminatory and biased features (Naik et al., 2022;

Mujtaba & Mahapatra, 2019). **DP10** envisions the inclusion of stakeholders such as developers, users, and government in the development and use of ethical AI. Through integration, human autonomy can be ensured (John-Mathews, 2022; Buenfil et al., 2019). Arrieta et al. (2020) divide stakeholders and their respective desiderata. Stakeholders' involvement can be achieved through questionnaires or interviews. **DP11** points to representing ethical AI as a functional tool. It can relieve work and increase the well-being of workers in a psychological sense. Decision support saves time and reduces errors. An objective view is possible, reducing and eliminating bias (Rakowski et al., 2021; Schrader & Ghosh, 2018). **DP12** mentions fact-checking by AI systems to uncover disinformation. Natural language is used to identify summaries of information about an author's ideological stances. Deep fakes must be used to detect fake content by analyzing the characteristics of a subject (Lange & Lechterman, 2021).

4.4. Prioritization of Design Principles

We derived DPs and categorized four identified regions, countries, and cultures to the DPs. We rely on our literature dataset, primarily **KT3**, and publications from the governments of India, China, the USA, and Western Europe on their AI projects. We focused on the descriptions of ethical AI. We assign priorities to DPs by region, country, and culture. Literature provides principles and guidelines for ethical AI that are general or limited to one country. Assigning priorities regarding DPs allows us to understand the development and focus of regions, countries, and cultures, see Table 2. Dark red stands for high priority, yellow for low, red and orange priorities in between. Each DP must be assigned four priorities, and equal priorities are excluded. For example, **DP1** has the highest priority in Western Europe, the second priority in the USA, the third priority in India, and the last priority in China. This ranking was made by the selected and used literature in our study.

For an overall evaluation, numbers were assigned to the prioritizations (yellow=4; orange=3; red=2; dark red=1) and subsequently added per country. The higher the value, the lower the prioritization. India has the lowest prioritization in our ranking (total: 58). China is in third place (44). Western Europe (23) and the USA (25) are strongly ahead, but the USA prioritizes most of our DPs. Robinson et al. (2020) examine the influence of cultural values on AI in Nordic countries under the cultural dimensions according to Hofstede (2023). They illustrate that those who use AI will be estranged if these people are not involved in AI implementation. Birhane et al. (2022) identify the weak thematization of social factors into ethical AI in research (**DP10**). Western Europe and the USA prioritize stakeholder integration, and India prioritizes stakeholder involvement. In China, stakeholder implementation for ethical AI is done equally low. China's government decides on the capabilities and policies of AI systems, disregarding societal opinions. Brendel et al. (2021) argued that ethical considerations are culturally shaped, and the importance of ethical considerations in AI systems should be emphasized. Due to the different cultural differences between the four regions, countries, and cultures, there are different priorities for our DPs. Based on our results, the connection of AI systems to the Internet is viewed critically in China (**DP13**), while Western Europe is less critical of the connection to the Internet, despite privacy policies and regulations. Another feature is the specification of the maximum required computing power (**DP15**). China is highly technology-supportive, striving for computing power. China prioritizes this DP more than Western Europe.

Table 2. Priorities of the DPs in terms of regions, countries, and cultures

DPs	India	China	Western Europe	USA
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
Sum				

5. Primary Adjustments

Adjusting the usability, comprehensibility, and applicability of our KTs, DRs, and DPs is a relevant step of the DSR-oriented approach (Gregor et al., 2020; Vom Brocke et al., 2020). We surveyed three experts. The experts were surveyed in written form. Table 3 provides the expert (E) profiles. The experts were selected based on their experience. More experts were asked to provide a statement; however, many were unable or refused to provide statements. The experts work in organizations that advertise the implementation and use of Ethics in AI systems. All experts have an education in AI integration, technology, or programming. The country distribution enables a different perspective based on the experts' cultural and technological advances. The initial

KT, DRs, and DPs and their descriptions were sent to the experts. The experts were asked to rate the usability, comprehensibility, and applicability of KTs, DRs, and DPs. New relationships created by adjustment are marked with blue arrows in the Figure in the online Appendix C, while DRs and DPs changed or added are shown in a dark gray with the signature "Adjustment."

Table 3. Expert profile

E	Description	Country
1	Inventor of SaaS Knowledge Graph that can analyze, infer, and chat considering ethical standards, rules, and norms	United Kingdom
2	Chairman of a company that ensures customers' AI systems operate equitably, ethically, and safety	USA
3	Chief ML Research Scientist	Germany

DP5 was supplemented in the call for accessibility and availability in all parts of society with higher equality (Expert 3). **DP2** was expanded to include systems that detect and prevent unintended damage or malfunction (Expert 1). **DP13** assumes that AI systems must not be connected to the Internet or other general relationships that allow AI systems to be hacked or perform offensive hacking (Expert 2, 3). **DP14** limits intelligence implementation in AI systems to task-related intelligence (Expert 2, 3). Necessary intelligence must be supplied to the system to accomplish the task. **DP15** requires ethical AI not to receive additional computing power beyond performing the maximum of their tasks (Expert 2). **DR13** describes the corrigibility of an ethical AI (Expert 2). The system must tolerate and support the programmer; it must not tamper; it must be able to repair safety measures; the programmer or user must be able to correct the system.

6. Discussion, Recommendations, Limitations, and Further Research

Regarding **DR5**, Expert 1 stated that "these requirements are subjective measures. AI or ML systems depend on their control source or dataset." **DP9** can support this statement. Therefore, we included access, availability, integration, and quality of data in **DP9**. Expert 1 highlighted the need to develop DPs continuously and identify trends from the literature. The experts acknowledge that security and protection need more research, encouraging companies to develop protection measures. Expert 2 stated that integrating protection measures "is only possible if we have access to cybersecurity and information protection experts." Expert 3 emphasized the value of human accountability, stating, "Humans must exercise judgment when using AI," which is consistent with our **DP1** showing a "clearly defined scope" (Expert 3) for AI systems. We see the difficulties of ethical AI as a functional tool. Because of decision-making, explanations of functional tools may have the weakest denotational power and thus satisfy the least desiderata of stakeholders (**DP10**). **DR5** calls for ethical AI to act non-discriminately, fairly, and diversely. Bias often originates in training datasets (Paraman & Anamalah, 2023; Mujtaba & Mahapatra, 2019). Teaching discriminating features to AI systems leads to non-ethical decisions by AI systems. Training datasets must avoid such a feature (**DP9**). It is important that data can be reviewed to avoid issues in data collection and mining (Expert 3). Another issue relates to **DP13**, as it is impossible to develop an AI system that is not connected to the Internet. While this protects the system and the data, it excludes **DP8**, **DP9**, and **DP15**.

Our assignment in Table 2 ranked the prioritization of our DPs by regions, countries, and cultures. India is the least advanced despite its progressive AI development and ethical standards implementation. This can be attributed to the general conditions in India. China ranks third in the prioritization of our DPs. Weaknesses can be attributed to the area of human supervision (**DP1**) and the implementation of data protection (**DP3**). China's prioritization of **DP13** is notable. China has developed initiatives for isolation with the Internet and AI systems. Western Europe connects most AI systems to the Internet but only uses the necessary intelligence in AI systems (**DP14**). Also notable in China is **DP15**, which deploys the maximum necessary computing power as the second prioritized country. The prioritization gap between Western Europe and the USA is small. This can be attributed to the subjective evaluation of the prioritization of our DPs.

We developed guidelines and principles for ethical AI, focusing on regions, countries, and cultures. We included and elaborated researched principles and guidelines (e.g., Bankins & Formosa, 2023; Nguyen et al., 2023; Paraman & Anamalah, 2023; Prem, 2023; Mikalef et al., 2022; Ryan & Stahl, 2021; Hagendorff, 2020; Peters et al., 2020; Floridi, 2019; Floridi & Cowls, 2019; Jobin et al., 2019; Yu et al., 2018). Most studies identify principles or guidelines that address six aspects: transparency, robustness and security, human oversight, privacy, community well-being, and accountability (e.g., Nguyen et al., 2023; Paraman & Anamalah, 2023; Floridi, 2019). Some studies mentioned more principles and guidelines (e.g., Bankins & Formosa, 2023; Prem, 2023; Ryan & Stahl, 2021; Jobin et al., 2019). Few publications identify country-specific principles and guidelines (e.g., Floridi & Cowls, 2019; Yu et al., 2018). We consider these principles and guidelines regarding the breadth of literature and solutions as requirements and add new and innovative principles. We create new arrows between existing results from the literature and our results. Hagendorff (2020) reviews guidelines and principles that relate to other countries. He notes that

only two of the identified publications in his study refer to cultural differences and emphases. In contrast to the other publications, we have been able to relate our DPs to four regions, countries, and cultures. We have expanded and received the DSR-oriented approach through DPs by prioritizing the DPs to four regions, countries, and cultures. We identified that India has the lowest prioritization rate. Researchers and organizations could fill the gap between these countries, cultures, and regions and derive guidelines for India or China. The USA and Western Europe are leading in ethical AI and can support India and China through ideas, research, and guidance. Including the literature in KTs allowed us to formulate specific DRs and DPs and build a structured research design. We created a knowledge base by developing a design artifact. Further research can expand our guidelines for implementing ethical AI. Our results and findings also strengthen the focus on ethical AI classifying and categorizing the ethical AI literature into KTs. We derived arrows between DRs and DPs limited in the literature. Our results and findings apply to multiple industries and sectors, e.g., energy and health. In contrast to other publications, e.g., Hagendorff (2020), we were able to guide the successful development and implementation of ethical AI through the deduced DPs. The derived DRs can be used to develop new solutions by researchers or organizations.

Due to the literature review, our analysis is limited to subjectivity. To reduce this limitation, all authors considered the publications separately. The clustering method was performed to achieve more objectivity. Another limitation is the general consideration of our DPs. Our results need further research in specific use cases. ***RQI)*** *"How can industry/sector-specific DPs for ethical AI be developed?"* We had our DPs adjusted by three experts from three different countries. For a general adjustment, further experts from different countries must be included to validate our DPs and the assignment of DPs to regions, countries, and cultures. Another RQ is ***RQII)*** *"How do experts from the USA, Western Europe, India, and China evaluate and adjust our DPs, and what prioritization of DPs do they suggest?"* Another research gap is the lack of literature on principles and guidelines for ethical AI from the perspective of different countries. The priorities of DPs need to be published by the countries and investigated in science. ***RQIII)*** *"How do design principles derive in different regions, countries, and cultures?"* We relied on normative Ethics based on our RQs and our goals. Other results regarding country selection and categorization of regions, countries, and cultures might emerge when considering other Ethics. Other Ethics could be Metaethics, applied Ethics, or Ethics in design (Brendel et al., 2021). ***RQIV)*** *"How can design principles for ethical AI derive when considering different Ethics?"*

7. Conclusions

To address our RQs, we followed a nine-step DSR-inspired approach based on Vom Brocke et al. (2020). We addressed RQ1 based on the cultural dimension according to Hofstede (2023; 2010). To address RQ2, we used text mining and topic modeling based on Gerlach et al. (2022) and Tong & Zhang (2019) and deduced four clusters that can be interpreted as KTs. Based on these and a systematic literature review, we derived 13 DRs and formulated 15 DPs. We adjusted our design artifact (DPs), surveying experts who offered ethical considerations in their AI systems. We have categorized the regions, countries, and cultures based on the cultural dimensions of our DPs and determined that the USA and Western Europe are more advanced in implementing and considering our DPs than India and China. Based on this, we discussed relationships. We provided a further research agenda, including RQs.

8. Acknowledgements

The research project “SiNED—Systemdienstleistungen für sichere Stromnetze in Zeiten fortschreitender Energiewende und digitaler Transformation” acknowledges the support of the Lower Saxony Ministry of Science and Culture through the 'Niedersächsisches Vorab' grant programme (grant ZN3563).

9. References

AI, H. (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html. Accessed: June 12, 2023.

Akata, Z., Balliet, D., et al. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect with Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer, 53*(8), 18–28.

Arrieta, A. B., Díaz-Rodríguez, N., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion, 58*, 82–115.

Bankins, S., & Formosa, P. (2023). The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. *Journal of Business Ethics*, *185*, 725-740.

Berente, N., Gu, B., Recker. J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly, 45*(3), 1433–1450.

Birhane, A., Ruane, E., Laurent, T., S. Brown, M., Flowers, J., Ventresque, A., & L. Dancy, C. (2022). The Forgotten Margins of AI Ethics. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*.

Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L. (2021). Ethical Management of Artificial Intelligence. *Sustainability*, *13*(4), 1974.

Buenfil, J., Arnold, R., Abruzzo, B., & Korpela, C. (2019). Artificial Intelligence Ethics: Governance through Social Media. *Proceedings of the 18th IEEE International Symposium on Technologies for Homeland Security*.

Carter, S. M., Rogers, W., Win, K. T., Frazer, H., Richards, B., & Houssami, N. (2020). The Ethical, Legal and Social Implications of Using Artificial Intelligence Systems Breast Cancer Care. *The Breast, 49*, 25–32.

Chao, C. H. (2019). Ethics Issues Artificial Intelligence, *Proceedings of the 24th IEEE International Conference on Technologies and Applications of Artificial Intelligence*.

Chatterjee, S., & NS, S. (2022). Artificial Intelligence and Human Rights: A Comprehensive Study from Indian Legal and Policy Perspective. *International Journal of Law and Management, 64*(1), 110-134.

Currie, G., Hawk, K. E., & Rohren, E. M. (2020). Ethical Principles for The Application of Artificial Intelligence (AI) Nuclear Medicine. *European Journal of Nuclear Medicine and Molecular Imaging, 47*(4), 748–752.

Demšar, J., Curk, T., Erjavec, A., et al. (2013). Orange: Data Mining Toolbox Python. *Journal of Machine Learning Research, 14*, 2349–2353.

Dolganova, O. I. (2021). Improving Customer Experience with Artificial Intelligence by Adhering to Ethical Principles. *Business Informatics, 15*, 34–46.

Eurostat (2023). Demographic Balances and Indicators by Type of Projection. https://ec.europa.eu/eurostat/databrowser/view/proj_23ndbi/default/table?lang=en

Feldman, M. D., Zhang. J., & Cummings, S. R. (1999). Chinese and US Internists Adheres to Different Ethical Standards. *Journal of General Internal Medicine, 14*, 469-473.

Floridi, L. (2019). Establishing the Rules for Building Trustworthy AI. *Nature Machine Intelligence, 1*(6). 261-262.

Floridi, L., & Cowls. J. (2019). A Unified Framework of Five Principles for AI Society. *Machine Learning and the City: Applications Architecture and Urban Design*, 535-545.

Ford, R., & Jennings, W. (2020). The Changing Cleavage Politics of Western Europe. *Annual Review of Political Science, 23*, 295-314.

Forsyth, S., Dalton, B., Foster, E. H., Walsh, B., Smilack. J., & Yeh, T. (2021). Imagine a More Ethical AI: Using Stories to Develop Teens Awareness and Understanding of Artificial Intelligence and its Societal Impacts, *Proceedings of the 6th IEEE Conference on Research Equitable and Sustained Participation Engineering, Computing, and Technology*.

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and Legal Challenges of Artificial Intelligence-driven Healthcare. *Artificial Intelligence Healthcare*, Elsevier, 295–336.

Gerlach, J., Scheunert, A., & Breitner, M. H. (2022). Personal Data Protection Rules! Guidelines for Privacy-Friendly Smart Energy Services. *Proceedings of the 30th European Conference on Information Systems*.

Ghotbi, N., Ho, M. T., & Mantello. P. (2022). An Attitude of College Students Towards Ethical Issues of Artificial Intelligence an International University Japan. *AI & Society, 37*(1), 283–290.

Green, B. P. (2018). Ethical Reflections on Artificial Intelligence. *Scientia et Fides, 6*(2), 9–31.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly, 37*(2), 37–355.

Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of Association for Information Systems, 21*(6), 1622–1652.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, 30*(1), 99-120.

Hevner, A. R., & Chatterjee, S. (2010). The Use of Focus Groups Design Science Research. *Design Science Research Information Systems: Theory and Practice*, Springer, 9–22.

Ho, J., & Wang, C. M. (2021). Human-Centered AI using Ethical Causality and Learning Representation for Multi-Agent Deep Reinforcement Learning. *Proceedings of the 2nd IEEE International Conference on Human-Machine Systems*.

Hofstede, G. (2010). Cultures and Organizations: Software of the Mind. McGraw-Hill.

Hofstede, G. (2023). Compare Countries – Hofstede Insights. https://www.hofstede-insights.com/product/compare-countries/. (Accessed: June 12, 2023).

Ingram, K. (2020). AI and Ethics: Shedding Light on the Black Box. *The International Review of Information Ethics, 28*.

Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Joh, E. E. (2022). Ethical AI American Policing. *Notre Dame Journal on Emerging Technologies*.

John-Mathews, J. M. (2022). Some Critical and Ethical Perspectives on the Empirical Turn of AI Interpretability. *Technological Forecasting and Social Change, 174*, #121209.

Kalyanakrishnan, S., Panicker, R. A., Natarajan, S., & Rao, S. (2018). Opportunities and Challenges for Artificial Intelligence India. *Proceedings of the 32nd AAAI/ACM Conference on AI, Ethics, and Society*.

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys, 55*(2), 1–38.

Lange, B., & Lechterman, T. M. (2021). Combating Disinformation with AI: Epistemic and Ethical Challenges. *Proceedings of the 49th IEEE International Symposium on Technology and Society.*

Leimanis, A., & Palkova, K. (2021). Ethical Guidelines for Artificial Intelligence Healthcare from the Sustainable Development Perspective. *European Journal of Spatial Development, 10*(1), 90–102.

Marda, V. (2018). Artificial Intelligence Policy India: A Framework for Engaging the Limits of Data-Driven Decision-Making. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), #20180087.

McGregor, C., Dewey, C., & Luan, R. (2021). Big Data and Artificial Intelligence Healthcare: Ethical and Social Implications of Neonatology. *Proceedings of the 49th IEEE International Symposium on Technology and Society*.

Meske, C., & Bunde, E. (2021). Transparency and Trust Human-AI-Interaction: The Role of Model-Agnostic Explanations Computer Vision-Based Decision Support. *Proceedings of the 22nd HCI International Conference*.

Mikalef, P., Conboy, K., Lundström. J. E., & Popovič, A. (2022). Thinking Responsibly About Responsible AI and the Dark Side of AI. *European Journal of Information Systems, 31*(3), 257-268.

Mirbabaie, M., Brendel, A. B., & Hofeditz, L. (2022). Ethics and AI Information Systems Research. *Communications of the Association for Information Systems, 50*(1), 38.

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical Considerations AI-Based Recruitment. *Proceedings of the 47th IEEE International Symposium on Technology and Society*.

Naik, N., Hameed, B. M. Z., et al. (2022). Legal and Ethical Consideration Artificial Intelligence Healthcare: Who Takes Responsibility? *Frontiers Surgery, 9*, #862322.

Nguyen, A., Nga, N. H., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical Principles for Artificial Intelligence in Education. *Education and Information Technologies*, *28*, 4221-2141.

Nwafor, I. E. (2021). AI Ethical Bias: A Case for AI Vigilantism (Ailantism) Shaping the Regulation of AI. *International Journal of Law and Information Technology, 29*(3), 225–240.

Paraman, P., & Anamalah, S. (2023). Ethical Artificial Intelligence Framework for a Good AI Society: Principles, Opportunities, and Perils. *AI & Society*, *38*, 595-611.

Pesapane, F., Volonté, C., Codari, M., & Sardanelli, F. (2018). Artificial Intelligence as a Medical Device Radiology: Ethical and Regulatory Issues Europe and the United States. *Insights into Imaging, 9*, 745-753.

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—Two Frameworks for Ethical Design Practice. *IEEE Transactions on Technology and Society, 1*(1), 34–47.

Prem, E. (2023). From Ethical AI Framework to Tools: A Review of Approaches. *AI & Ethics*, *3*, 699-716.

Rakowski, R., Polak. P., & Kowalikova. P. (2021). Ethical Aspects of the Impact of AI: The Status of Humans the Era of Artificial Intelligence. *Society* 58(3), 196–203.

Roberts, H., Cowls. J., Hine, E., Morley. J., Wang, V., Taddeo, M., & Floridi, L. (2022). Governing Artificial Intelligence China and the European Union: Comparing Aims and Promoting Ethical Outcomes. *The Information Society*, *39*(2), 79-97.

Robinson, S. C. (2020). Trust, Transparency, and Openness: How Inclusion of Cultural Values Shapes Nordic National Public Policy Strategies for Artificial Intelligence (AI). *Technology in Society*, *63*, #101421.

Ryan, M., & Stahl, B. C. (2021). Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications. *Journal of Information, Communication and Ethics Society, 19*(1), 61–86.

Schrader, D. E., & Ghosh, D. (2018). Proactively Protecting Against the Singularity: Ethical Decision-Making AI. *IEEE Security & Privacy, 16*(3), 56–63.

Seo, H., & Thorson, S. (2022). Computation, Rule Following, and Ethics AIs. *Proceedings of the 55th Hawaii International Conference on System Sciences*.

Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting Fundamental Rights and Ethical Values. *Computer Law & Security Review*, 45, #105661.

Steimers, A., & Bömer, T. (2021). Sources of Risk and Design Principles of Trustworthy Artificial Intelligence. *Proceedings of the 12th International Conference on Digital Human Modeling and Applications Health, Safety, Ergonomics and Risk Management*.

Tong, Z., & Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. *Proceedings of the 3rd International Conference on Computer Science, Engineering and Information Technology*, 201–210.

United Nations (2022). World Population Prospects 2022: Summary of Results. *Department of Economic and Social Affairs.*

Vallor, S. (2016). Technology and the virtues: A Philosophical Guide to a Future Worth Wanting. *Oxford University Press*.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search Information Systems Research. *Communications of the Association for Information Systems, 37*(9), 205–224.

Vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special Issue Editorial–Accumulation and Evolution of Design Knowledge Design Science Research: A Journey Through Time and Space. *Journal of the Association for information Systems, 21*(3), 520–544.

Wang, Z., Tang, C., Sima, X., & Zhang, L. (2020). Research on Ethical Issues of Artificial Intelligence Technology. *Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture*.

Watson, R. T., & Webster. J. (2020). Analysing the Past to Prepare for the Future: Writing a Literature Review a Roadmap for Release 2.0. *Journal of Decision Systems, 29*(3), 129-147.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly, 26*(2), xiii–xxiii.

Wu, W., Huang, T., & Gong, K. (2020). Ethical Principles and Governance Technology Development of AI China. *Engineering, 6*(3), 302-309.

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., & Yang, Q. (2018). Building Ethics into Artificial Intelligence. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*.

Zhang, J., Shu, Y., & Yu, H. (2023). Fairness in Design: A Framework for Facilitating Ethical Artificial Intelligence Designs. *International Journal of Crowd Science*, *7*(1), 32-39.