


Bridging East-West Differences in Ethics Guidance for AI and Robotics

Nancy S. Jecker^{1,2,3}  and Eisuke Nakazawa^{1,4,*}

¹ Department of Bioethics and Humanities, University of Washington School of Medicine, Seattle, WA 98195-7120, USA

² Centre for Bioethics, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

³ Department of Philosophy, University of Johannesburg, Johannesburg 2006, South Africa

⁴ Department of Biomedical Ethics, University of Tokyo Graduate School of Medicine, Tokyo 113-0033, Japan

* Correspondence: nakazawa@m.u-tokyo.ac.jp; Tel.: +81-35841-3511; Fax: +81-35841-3319

Abstract: Societies of the East are often contrasted with those of the West in their stances toward technology. This paper explores these perceived differences in the context of international ethics guidance for artificial intelligence (AI) and robotics. Japan serves as an example of the East, while Europe and North America serve as examples of the West. The paper's principal aim is to demonstrate that Western values predominate in international ethics guidance and that Japanese values serve as a much-needed corrective. We recommend a hybrid approach that is more inclusive and truly 'international'. Following an introduction, the paper examines distinct stances toward robots that emerged in the West and Japan, respectively, during the aftermath of the Second World War, reflecting history and popular culture, socio-economic conditions, and religious worldviews. It shows how international ethics guidelines reflect these disparate stances, drawing on a 2019 scoping review that examined 84 international AI ethics documents. These documents are heavily skewed toward precautionary values associated with the West and cite the optimistic values associated with Japan less frequently. Drawing insights from Japan's so-called 'moonshot goals', the paper fleshes out Japanese values in greater detail and shows how to incorporate them more effectively in international ethics guidelines for AI and robotics.

Keywords: artificial intelligence; ethics; robotics; Japan; philosophy of technology; international ethics



Citation: Jecker, N.S.; Nakazawa, E. Bridging East-West Differences in Ethics Guidance for AI and Robotics. *AI* **2022**, *3*, 764–777. <https://doi.org/10.3390/ai3030045>

Academic Editors: Pablo Rivas, Gissella Bejarano and Javier Orduz

Received: 27 July 2022

Accepted: 7 September 2022

Published: 14 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Societies of the East are often contrasted with those of the West in their stances toward technology. This paper explores these perceived differences in the context of international ethics guidance for artificial intelligence (AI) and robotics. Its principal aim is bridging these differences and arriving at a more inclusive, truly 'international' ethics guidance. The paper also contributes to 'cultural robotics', or the study of cultural values embedded in robotic design [1].

Japan serves as an example of the societies of the East, while Europe and North America serve as examples of the West. We select these societies because we are most familiar with them and because they have well-developed robot cultures traceable to the aftermath of the Second World War. While we cite evidence suggesting that values we identify may hold true for other societies of the East and West, we bracket a fuller discussion for another day. Throughout the paper, we name certain traits "Japanese" and others "Western" to indicate that they are widely found among people in those regions. We do not mean to imply that all or only people in those regions hold the views in question or that they are untouched by outside influences.

Sections 2 and 3 examine how distinct stances toward robots emerged in Japan and the West during the aftermath of the Second World War, reflecting history and popular culture, socio-economic conditions, and religious and philosophical worldviews. While a

hallmark of Japan's approach is technological optimism, the West reflects a more precautionary stance. Section 4 considers how these distinct stances of optimism and precaution appear in international ethics guidelines for AI and robotics, drawing on a 2019 scoping review that examines 84 international AI ethics documents. These documents are heavily skewed toward precautionary values associated with the West and cite the optimistic values associated with Japan less frequently. Drawing insights from Japan's *Moonshot Goals for a Robot-Human Co-existence Society*, the paper fleshes out the six less cited values and recommends a hybrid approach that better incorporates them in international ethics guidance for AI and robotics.

2. Western Views toward Robots

In a 2019 scoping review, Jobin et al. identified 84 documents from around the world containing AI ethics guidance, mostly produced by private companies and governmental agencies [2]. Although no principles were common to each, 11 values stood out as common to many. Of these, five were endorsed in the majority of the documents.

Five Majority Values

- (i). Transparency;
- (ii). Justice/fairness;
- (iii). Non-maleficence;
- (iv). Responsibility;
- (v). Privacy.

These five values were characterized as precautionary, stressing preventing harm, and minimizing risk. The six remaining values occurred in a minority of the AI ethics documents.

Six Minority Values

- (i). Beneficence;
- (ii). Freedom and autonomy;
- (iii). Non-maleficence;
- (iv). Sustainability;
- (v). Dignity;
- (vi). Solidarity.

These values were characterized in positive ways, emphasizing how technology can contribute to a good society and flourishing human lives. However, the positive values were less frequently cited and less fully elaborated when they were identified. For this reason, Jobin et al. characterize the five majority values as revealing a "negativity bias" in international AI ethics guidelines. They go on to observe that "Because references to non-maleficence outnumber those related to beneficence, it appears that issuers of guidelines are preoccupied with the moral obligation to prevent harm" [2].

This section will show that the majority of negative values listed here (and further discussed in Section 4) align with values and attitudes toward robots prevalent in the history, popular culture, socio-economic policies, and religious and philosophical worldviews of the West, indicating that what Jobin et al. call a 'negativity bias' may also reflect a Western bias. Since international AI ethics guidance aspires to respond to ethical problems across geographies and cultures, the broader concern Jobin et al.'s study raises is not just that ethics guidance is predominantly precautionary, but that Western values are put forth as universally applicable. This can result in "little effort to discover, teach, and integrate . . . other moral traditions" [3]. Bamford expresses concern that universal ethical approaches "may not be sufficiently inclusive of cultural diversity and may foster cultural imperialism" [4]. Davis puts it thus: "Most Western philosophers continue either to ignore other traditions of thought or, when they do turn their attention to them, to approach them more or less monologically" [5]. Fox depicts Western ethics as having a "molding" influence on other nations [6]. To avoid slipping into "unintended imperialism", finding and correcting Western bias is critically important [7].

2.1. The History of Robots in the West

Robots were not originally conceived in a laboratory but in a work of fiction. The Czech playwright, Karel Čapek, dreamed up the idea and coined the term *robot* to refer to it in the 1921 play *Rossum's Universal Robots (R.U.R.)* [8]. The play is a dystopian tale of robots mass produced by other robots on assembly lines. The plot features a character who begins to believe that robots have (or are developing) souls and should be freed. They press to enhance robots to enable them to develop souls more quickly and fully. The plan ends in disaster with an enhanced robot ordering the robots of the world to exterminate humanity while sparing machines: "Robots of the world, we enjoin you to exterminate mankind. Don't spare the men. Don't spare the women. Retain all factories, railway lines, machines and equipment, mines, and raw materials. All else should be destroyed" [8].

R.U.R. sparked debate over whether human-made machines could ever develop souls or other qualities, such as thinking and emotions, previously considered uniquely human. Simultaneously, it raised existential questions about whether robots would replace or destroy humans if they learned to 'make themselves'. When *R.U.R.* was first performed in English in 1922, the English word, *robot*, appeared for the first time, derived from the Czech, *robot*, which means "forced labour, hard work", and has multiple origins, including "slavery" (from Church Slavic, *rabota*, *robot*) and "servant" (from Old Russian, *rab*); its historical meaning was "a central European system of serfdom, in which a tenant's rent was paid in forced labour or service" [9].

The work robots were forced to perform in *R.U.R.* reflected the socio-economic conditions of people's lives during the early industrial revolution in Europe and North America. During this time, factories replaced the domestic system of production, which relied on handmade goods, with mass-produced machine-made products. While assembly lines increased efficiency and made more goods available at a lower cost, they resulted in people working long hours for low wages, and being exposed to unsafe conditions, which, at the time, they had little legal standing to challenge. Under these circumstances, "people began to wonder if they were being replaced by machines, or even becoming machines themselves" [10]. Popular fiction reflected their angst. It spoke to the perception that robots were threats to a way of life, even when they created cheaper and more abundant goods.

2.2. Robots Reflected in Popular Culture, Socio-Economic Policies, and Religious and Philosophical Worldviews of the West

The original perception that robots represent a danger and threat is a theme that reverberates throughout fiction and film in the West during the post-World War II era. A variation of it appeared in the 1927 German film, *Metropolis* (based on a 1925 novel of the same name), which told the story of a gleaming city powered by underground-dwelling workers, who lived in slavish conditions operating machines that kept the wealthy, above-ground city going. Similar to *R.U.R.*, *Metropolis* ends in chaos and destruction, wrought by a deceptive robot who impersonates a human being.

Chaos and destruction are also apparent in more recent film and fiction emanating from the West. For example, the 1968 classic, *2001: A Space Odyssey*, features Hal, a mainframe computer, taking over a spaceship, turning off oxygen supplies, and killing virtually everyone on board. In the 1973 film, *Westworld*, android entertainers at a futuristic amusement park go haywire and start killing visitors. During the 1980s, *Terminator* (1984) depicted a killer cyborg that is stronger and more powerful than humans, and *Blade Runner* (1982) showcased robots that were virtually indistinguishable from human beings and had souls, raising the question of which type of soul was higher—human or robot. *Matrix* (1999) put giant robots at the top of the food chain and portrayed humans as their battery sources. Dangerous robots were also on display in *Ex Machina* (2014), which featured a beautiful female robot, Ava, who feigns romantic interest in a human man to secure her freedom, leaving the human who loved her to die.

Alongside these are sprinkled more positive images. For example, the 1977 film series *Star Wars* featured R2-D2 and his sidekick, C-3PO, as friendly android companions to

Luke Skywalker. The gentle android, Commander Data, in the television series, *Star Trek the Next Generation* (1987), served as a loyal member of the starship's crew. The film, *A.I. Artificial Intelligence* (2001), depicted an endearing robot boy's attempt to become fully human. WALL-E (2008) puts on the big screen a friendly non-humanoid robot living on an environmentally devastated earth, suggesting that humans, not machines, are the menace who ruined the planet and made it uninhabitable for themselves. Still, the predominant perception one gleans from film and fiction in the aftermath of the Second World War is that robots pose a threat and danger to human beings.

When contemporary Western philosophers consider robots, their approaches echo not only concerns raised in film and fiction but also the wider Judeo-Christian worldview prevalent throughout the West. For example, it is assumed that robots are of a lower order and lack the inherent worth and dignity associated with human beings [11]. It is also assumed that robots are incapable of developing conscious states or subjective feelings of any kind. It stands to reason that if humans enter into close relationships with robots, they must be duped, erroneously thinking that sophisticated robots are like them. For example, Elder writes that robots deceive others with false appearances of friendship [12]. According to Turkle, we are changed for the worse as technology offers us "substitutes for connecting with each other face-to-face" [13] (p. 11). Sparrow laments that, "Despite their animated appearance, robots remain essentially inanimate objects. They can contribute nothing to the relationships that people form with them. The range of emotions appropriate toward a robot is thus limited to those that would be appropriate toward a car, wristwatch, or antique *settee*" [14] (p. 315).

According to this way of thinking, robots' proper role is to serve as instruments to achieve human goals. What Heidegger dubbed, "the instrumental definition of technology", captures this idea [15] (p. 5). It holds that human-machine relationships are always a means to some human end. This idea is apparent in a 2007 article by Microsoft founder, Bill Gates, entitled, "A Robot in Every Home", which envisioned the many tasks that robots would one day perform to reduce human drudgery: floor-cleaning, laundry-folding, lawn mowing, food and medicine dispensing, and surveilling [16]. In Gates' robotic future, robots do not figure in human social relationships as partners, companions, or friends, but only as objects working for human masters.

A similar view of robots as tools to achieve human objectives is evident in the deployment of robots during warfare. Revealingly, the limits of a purely instrumental view of robots also come to light in that setting. During wars in Afghanistan and Iraq, Packbots, a creation of iRobot, were substituted for human soldiers to defuse innovative electronic devices. Human soldiers working alongside Packbots eventually established close relationships with them, reportedly awarding robots medals and battlefield promotions; if they were severely damaged and 'died', soldiers held funerals [17]. Just as the military creed, 'leave no man behind', was applied to human soldiers, the same ethic was reportedly applied to robot soldiers: "... in Iraq, [a] soldier ran fifty meters, all the while being shot at by an enemy machine gun, to 'rescue' a robot soldier" [17] (p. 339). These responses suggest, "Our machine creations are not just 'neutral' objects to us. We not only tend to view them as having their own personalities, but also feel that they deserve some form of emotional attention and engagement" [17] (p. 340).

One explanation for human soldiers' responses, offered by Singer, is that "we are hard wired that way" [17] (p. 340). However, even if humans are 'hard wired that way', it remains an open question how we *ought* to respond to robots. Does it enhance human society to follow Sparrow's admonition to treat robots like cars or wristwatches, or is society enhanced by bonding with robots, as American soldiers reportedly did during the wars in Afghanistan and Iraq? In the West, with a few exceptions [18–21], philosophers recommend keeping emotional distance from robots. However, other societies, such as Japan, answer this question in a strikingly different way.

3. Japanese Views toward Robots

We turn next to the six more positive values reported by Jobin et al.: beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity. In this section, we show that these affirmative values align with the stances toward robots prevalent in the history, popular culture, socio-economic conditions, and religious and philosophical worldviews of Japan. Later in the paper (Section 4), we will say more about these positive values, elaborating them in greater detail by borrowing insights from Japan.

3.1. *The History of Robots in Japan*

Japan has a specific robot culture traceable to the aftermath of the Second World War, with roots in popular manga (comics and graphic novels) and anime (hand-drawn computer animation). In 1951, after the U.S. bombing of Hiroshima and Nagasaki and Japan's wartime defeat, manga artist, Osamu Tezuka, introduced *Testuwan Atom* or 'Mighty Atom' (known in the West as 'AstroBoy'), a friendly young android boy with an atomic heart, capable of defending Japan against threats from outer space. Mighty Atom's nuclear heart was reminiscent of the atomic bomb, and Tezuka explains the post-war backdrop for his creation thus: "I realized very clearly that Japan lost the war because of science and technology . . . While the U.S. was dropping atomic bombs, the Japanese military were trying to light forest fires in America by sending incendiary balloons made of bamboo and paper over the jet streams. We developed an inferiority complex about science" [22] (Kindle location 1561–1572). Tezuka, a physician, portrayed Mighty Atom's special powers as scientifically based: he was propelled by a 100,000 hp atomic engine, solved mathematical problems using a computer brain in his chest, sported a translation device in his throat enabling him to speak sixty languages, had eyes capable of functioning as search lights, and came equipped with a button on his ear to increase hearing capacity 1000 times. To battle bigger robots, Mighty Atom employed twin machine guns on his fanny. Mighty Atom was enormously popular, considered "a symbol of state-of-the-art Japanese technology" [22] (Kindle location 1561).

The character's 'robot nature' was "one of his most endearing charms"—he was humanlike, and strove to be more human, yet was clearly different [22] (Kindle location 1629). For example, absent special adjustments, Mighty Atom was unable to cry, did not usually experience fear, and could not feel true love. However, despite this, Schodt argues that Mighty Atom could clearly pass a Turing test (i.e., display the ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human), noting that in the manga plots in which he appears, he sometimes fools others into believing he is human [22] (Kindle location 1627).

The development of Mighty Atom was shaped, in part, by popular culture in the West. *R.U.R.* was performed in Tokyo in 1924 as *Jinzô Ningen (Artificial Human)*, and three years later, *Metropolis* appeared on Japan's big screen. Interviews with Tezuka reveal that, in creating Mighty Atom, these dystopian stories influenced him; however, rather than imitating them, Tezuka instead made Mighty Atom a positive figure who bridged two cultures—the human and machine [22] (Kindle location 1688). Not only did the robot boy stand down outside threats and protect Japan, but he abided by Asimov's Three Laws of Robotics—not harming humans, obeying human orders, and protecting humans from threats.

3.2. *Robots Reflected in Popular Culture, Socio-Economic Policies, and Religious and Philosophical Worldviews of Japan*

The perception of robots as a positive force and protector of society is a theme that reverberates in other Japanese manga and anime from the post-war era. For example, Mistuteru Yokoyama's manga character, *Tetsujin 28-go*, created in 1958, was a giant robot remotely controlled by a young boy. His technology saved Japan from an attacking enemy. Likewise, Doraemon, a cuddly blue robotic cat who appeared in 1970, traveled back in time to help ancestors and thereby improve the conditions for descendants. Contemporary

Japanese robots, such as Sony's robotic dog AIBO (released in 1999 and re-released in 2019), also exemplify a friendly robot culture. In Japanese, *aibō* means 'pal' or 'partner', and its creator, Toshitada Doi, designed AIBO to be a beloved family member who was adopted as a newborn puppy and developed into an adult dog with a personality shaped by owners. The close bond owners developed toward AIBO were apparent in 2014, when Sony stopped making replacement parts. This led owners to seek the services of A-FUN, a Japanese company specializing in repair of defunct electronics; A-FUN made unrepairable AIBOs into 'organ donors' for other AIBOs and commemorated the donors at funerals held at a 450-year-old Buddhist temple, Kōfuku-ji [23].

In addition to positive images, one finds occasional mixed messages about robots in Japan. For example, *Tetsujin 28-go*'s backstory describes him as a product of World War II, originally designed to wage war; however, he finds a new purpose in post-war Japan, as "the iron golem [used] not to wage war, but to protect peace and stop crime ... a weapon of destruction finding a new identity as a guardian of good ... " [24]. The underlying suggestion is that whether *Tetsujin 28-go* promotes good or evil depends on who operates the hand-held controller. In contrast, Atom Boy is an autonomous robot with a moral conscience. Tezuka "worked hard to counter the 'evil' image of metal men in the West, and as a result his creation—Mighty Atom—became a friend of man" [22] (Kindle location 1682).

In addition to popular culture, socio-economic circumstances shaped Japan's optimistic stance toward robots. The oldest society in the world, Japan, has the highest old-age dependency ratio of any nation and faces looming labor shortages [25] (p. 12). Robots are considered integral to solving Japan's labor crisis, not only filling gaps in the paid labor force but also relieving women of caregiving and domestic chores assigned to them, enabling more women to join the paid labor force [26]. Other socio-economic challenges robots and AI are expected to help Japan address include limiting climate change through renewable energies and improving access to infrastructure for remote populations [27]. In these respects, Japanese people consider robots allies helping solve social problems.

A positive outlook toward robots is also evident in the Japanese government's blueprint, "Society 5.0" (also called "the Super-Smart Society"), introduced in 2016 by former Prime Minister Shinzo Abe. 'Society 5.0' refers to a fifth industrial revolution, following Society 1.0 (hunter-gatherer), 2.0 (agricultural), 3.0 (industrialized), and 4.0 (information). Japan's Council for Science, Technology and Innovation defines this Super Smart Society as "A human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space" [28] (p. 1). According to the Council's vision, "In Society 5.0, ... people, things, and systems are all connected in cyberspace and optimal results obtained by AI exceeding the capabilities of humans are fed back to physical space. This process brings new value to industry and society" [29] (p. 2).

The future society will initially utilize AI and medical elderly care robots to perform tasks, such as "walking assistance, supervision and support through conversation", followed by "robots which understand a person's intentions", and ultimately, personal robots that are general purpose and "utilized as family members in daily life, solving the problem of nursing care and allowing people to live in peace" [29] (p. 7). While robots will initially cooperate with humans on the job, ultimately, they will solve the problem of a shrinking workforce by performing multiple functions autonomously and cooperating with each other; they will not only fill jobs but also enhance products by hyper-customizing them [29] (p. 6).

These descriptions suggest that rather than envisioning robots as an outside threat, Society 5.0 sees physical and cyber systems as converging. Unlike their Western counterparts, which mark clear boundaries between humans and robots, Japan's approach is more apt to regard humans and robots as a harmonious whole. Although Society 5.0 is a blueprint for Japanese society, it is simultaneously set forth to align with global goals, such as the United Nations' Sustainable Development Goals. The hope is that Society 5.0 can eventually guide

other societies facing similar challenges related to population aging, labor supply shortages, renewable energy, and environmental protection.

Legacies of animism shed further light on Japanese stances toward technology. While many Japanese consider themselves non-religious [30], Japanese views of religion are complex, emphasizing ritual practice rather than doctrinal belief, and moving fluidly between multiple traditions—birth is celebrated with Shintoism, marriage performed with Christianity, and death mourned with Buddhism [31]. Arguably, a dichotomy between ‘secular’ and ‘religious’ did not exist in Japan before the Meiji era (1868–1912) [30,32]; the ubiquity of gods in ordinary objects reflects this. The Japanese saying, *Yoyorozu-no-Kami* (8 million gods), expresses the idea that the Japanese see gods in everything. Gods are thought to dwell in innumerable natural objects. For example, mountains have mountain gods, and rice paddies have rice paddy gods; gigantic stones and trees are also believed to be inhabited by deities, as are objects of worship. Gods also dwell in everyday artifacts, such as sewing needles.

Applied to machines, animism lends itself to regarding robots as animated, too. Mori, a Japanese roboticist, characterizes a Buddhist outlook on robots this way, “There is buddha-nature in dogs and in bear, in insects and in bacteria. There must also be buddha-nature in the machines and robots that my colleagues and I make;” Mori adds, “by recognizing in machines and robots the same buddha-nature that pervades [one’s] own inner self . . . harmony between human beings and machines is achieved” [33] (pp. 174, 178–179).

Shintōism sheds further light on Japanese attitudes toward technology. In its modern form, Shintō is sometimes regarded as Japan’s native religion. However, Kuroda et al. stress the secularity of Shintō, describing it as the character of the Japanese people embodied in practices that integrate and join diverse elements [34]. According to Kuroda, at any given time in Japan’s history, “what constituted the religion and thought of the Japanese people was something . . . assimilated or formulated or fabricated by the people, whether it was native or foreign in origin” [34] (pp. 5, 20). This rendering of Shintō depicts it as a bricolage of multiple traditions animated by a variety of beings (machine/organic/human) interwoven from diverse elements.

Applied to robots, a Shintō approach can be rendered as integrative, joining human beings with AI systems. A proclivity to join varied elements is apparent, for example, in Japanese conceptions of robot-human relationships. As Kaplan notes, “At the end of the Second World War, one could have expected that nuclear energy would be associated by Japan with death and defeat. But instead of being diabolized, the destructive energy was reintegrated into fiction as a positive life principle”, instantiated in anime and manga, such as *Atom Boy* and *Tetsujin 28-go* [35]. These robotic figures are ‘Japanese’ in the sense that they join machines and humans together in a positive way that is both non-threatening and aesthetically pleasing, reflecting an aesthetic that prizes *kawaii* (cute infant-like objects). By rendering robots sweet and appealing, Japanese designers avoid Mori’s ‘uncanny valley’, which refers to the profound unease alleged to arise when robots almost, but not fully, resemble humans in appearance and movement, creating an ‘eerie’ or ‘uncanny’ feeling (*bukimi no tani*) [36].

In contrast to Western stances that perceive intelligent machines with trepidation, Japanese narratives display unbridled enthusiasm, depicting robot-human relationships as positive, helpful, and empowering. Unlike Judeo-Christian traditions, which place humans at the apex of a hierarchy and find robots a potential threat to humanity’s elevated rank, Shintō does not emphasize high and low, or separate spirit and matter; instead, it seeks to connect distinct elements [37]. Sometimes referred to as ‘techno-animism’ [38], Japanese-inspired technology strives to bring humans and nonhumans closer together, breaking down barriers that divide them.

Today, positive images of robots abound in Japan, often reminiscent of the country’s early robot culture. Toyota’s 2013 companion robot, Kirobo Mini, was modeled after *Mighty Atom*; Sony introduced its companion robot, QRIO, in an episode of a 2003 *Tetsuwan Atomu*

anime; and Honda promoted ASIMO rentals on Atom's 'birthday'. Such efforts convey a deliberate attempt to promote a positive robot narrative [39].

Japanese/Western variances continue to appear today in public perceptions of robots. For example, a 2020 PEW Research Center study reported that the majority of most Asia-Pacific publics surveyed considered the effect of AI on society to be positive, while in places such as the Netherlands, the UK, Canada, and the U.S., publics are less enthusiastic and more divided on this issue [40]. At least two-thirds of survey respondents in five of the six Asian publics surveyed said that "computer systems designed to imitate human behaviors" had been good for society: Singapore (72%), South Korea (69%), India (67%), Taiwan (66%), and Japan (65%) (Funk et al., 2020). By contrast, in the U.S., less than half (47%) of respondents believe that the development of AI has been good for society, and fewer still (41%) think that using robots to automate jobs was a good thing.

4. Japanese and Western Views in International Ethics Guidelines

Stepping back from the analyses of Sections 2 and 3, this section considers the challenge of designing international ethics guidelines for AI and robotics in the face of dissimilar values and stances toward technology. In the case of Japanese–Western variances, at least three responses are possible. First, ethics guidance might reflect the belief that Western stances are roughly right: we should err on the side of safety and proceed with caution to avert worst-case scenarios. In other words, when developing new types of machines and AI systems, we should say 'better safe than sorry'. Second, it might be thought that a Japanese-like approach should guide. This response portrays robot-human relationships as positive and seeks to establish robots as potential allies in creating better societies and more flourishing human lives. A third approach is hybrid, featuring Western and Japanese stances as complementary, each valid yet incomplete. In this section, we illustrate these views and defend a hybrid approach.

To focus our discussion, consider again the 2019 scoping review by Jobin et al. (introduced in Section 2), which identified five majority values emphasizing precaution and avoidance of harm (transparency, justice/fairness, non-maleficence, responsibility, and privacy) and six minority values stressing positive benefits (beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity).

4.1. Precautionary Values

To further elaborate, consider first the five values most frequently adopted in international ethics guidelines (summarized in Table 1). We argued (in Section 2) that these values align closely with the history, popular culture, socio-economic conditions, and religious and philosophical worldviews of the West.

Table 1. Precautionary Values in International Ethics Guidance.

| | Values | Harms Avoided | Elaboration |
|-----|------------------|-----------------------------|--|
| (1) | Transparency | Deception and manipulation | Disclosing automation and data usage |
| (2) | Justice/Fairness | Implicit bias | Reducing algorithmic bias |
| (3) | Non-maleficence | Physical/psychological harm | Preventing cyberwarfare or malicious hacking |
| (4) | Responsibility | Non-accountability | Reforming liability laws, whistleblowing |
| (5) | Privacy | Privacy violations | Certificates of authenticity, data minimization, educating consumers |

These five values are oriented primarily toward preventing potential harms. *Transparency* speaks to concerns that robots are potentially manipulating, misleading, and deceiving users, and seeks to prevent this by disclosing data usage, automating decisions, and controlling the purpose to which users' data will be put. *Justice* encompasses efforts to minimize unwanted bias and discrimination and to offer redress when it occurs. *Non-maleficence* indicates risk management and minimizing intentional misuse. *Responsibility and accountability* address holding AI systems accountable for potential misdeeds and avoiding unwanted conduct using processes capable of preventing them. Finally, *privacy* stress protecting data from outside threats.

An often-cited formulation of this approach is a precautionary principle. While variously defined, a precautionary principle generally holds that precaution should inform public policy if a new technology has potentially devastating consequences, even if there is a high degree of scientific uncertainty about whether these consequences will occur [41]. In its most extreme form, precaution has been interpreted as anti-science; however, it is usually understood as a requirement to take reasonable precautions to deal with plausible threats [42]. For example, Hughes suggests the following formulation: "If there is evidence stronger than *E* that an activity will cause harm more seriously than *S*, perform action type *A* as a precaution against *S*" [43] (p. 452). Broadly understood, a precautionary principle manages risk using a 'safety first' strategy. It aligns with Western approaches to robots and AI, which emphasize transparency to avoid deception and manipulation, fairness to avoid algorithmic bias, nonmaleficence, legal accountability, and privacy protection.

A negative slant of the kind emphasized in most international AI ethics documents can have the unfortunate effect of sidelining positive values and downplaying affirmative possibilities, resulting in a one-sided analysis. It can neglect consideration of how AI systems can contribute to flourishing human lives and a good society. Considering both upsides and downsides of risk-taking is necessary for a balanced view, one that sometimes supports avoiding risks, and other times supports risk-taking.

4.2. Positive Values

In Jobin et al.'s scoping review, positive values included doing good or beneficence, enhancing freedom and autonomy, building trust, ensuring sustainability, affirming human dignity, and strengthening solidarity. We argued (in Section 3) that these values align closely with the history, popular culture, socio-economic conditions, and religious and philosophical worldviews associated with Japan. They were less frequently cited in international AI ethics guidance, and when cited, they were often not elaborated. Insights from Japan can potentially help flesh out these neglected values.

Consider, for example, Japan's current vision for a positive human-robot co-existence society. In 2018, Japan's Council for Science, Technology, and Innovation launched a "Moonshot Research & Development Program" designed to realize a technology-integrated society through "disruptive innovations" that go beyond existing technologies in their approach to robot-human co-existence. The program features seven moonshot goals, summarized in Table 2, with funding for associated research and development projects [44] (p. 12).

Moonshot Goal 1 calls for creating a society in which human beings can be free from limitations of body, brain, space and time. This goal funds research such as Kanai Ryota's, "Liberation from Biological Limitations via Physical, Cognitive and Perceptual Augmentation", which establishes an avatar-symbiotic society that uses cybernetic avatars that can be controlled via intention.

Moonshot Goal 2 involves ultra-early disease prediction and intervention. It underwrites research such as Katagiri Hideki's project to implement strategies to easily detect subjects in pre-symptomatic states of diabetes and comorbidities to prevent the development of these diseases.

Table 2. Japan’s Moonshot Goals.

| |
|--|
| MOONSHOT GOAL 1: a society in which human beings can be free from limitations of body, brain, space, and time; |
| MOONSHOT GOAL 2: ultra-early disease prediction and intervention; |
| MOONSHOT GOAL 3: AI-equipped robots that autonomously learn, adapt to their environment, evolve in intelligence, and act alongside human beings; |
| MOONSHOT GOAL 4: sustainable resource circulation to recover the global environment; |
| MOONSHOT GOAL 5: industry that enables sustainable global food supply by exploiting unused biological resources; |
| MOONSHOT GOAL 6: a fault-tolerant universal quantum computer that will revolutionize the economy, industry, and security; and |
| MOONSHOT GOAL 7: sustainable care systems to overcome major diseases and enable enjoying one’s life with relief from health concerns. |

Moonshot Goal 3 is designing AI-enabled robots that autonomously learn, adapt to their environment, evolve in intelligence, and act alongside human beings. It supports research to make robots that “humans feel comfortable with, have physical abilities equivalent to or greater than humans, and grow in harmony with human life” [44] (p. 7). An example is Hirata Yshuhisa’s project, “Adaptable AI-Enabled Robots to Create a Vibrant Society”, which aims to deploy adaptable AI-enabled robots at a variety of places usable by anyone at any time that will “adjust its form and functions according to the individual user to provide optimal assistance and services” [44] (p. 7).

Moonshot Goal 4 consists of developing sustainable resource circulation to recover the global environment. It targets innovative technologies, such as Direct Air Capture, which aims to “directly capture CO₂ that has already been released into the atmosphere and utilize it effectively” [44] (p. 8). Other projects aspire to design technologies that can detoxify nitrogen compounds released into the environment and convert them into valuable materials, and technologies to create biodegradable plastics replete with “degradation initiation switches” that are safe for the environment.

Moonshot Goal 5 is creating technologies to enable a sustainable global food supply by exploiting unused biological resources. This goal funds 10 research and development projects. Among them is a project to digitally design crops that can be grown in extreme environments, a project to raise cows with less methane emissions, and a project to convert wood materials left unused on forest floors into food and feed with the help of termites.

Moonshot Goal 6 consists of building a fault-tolerant universal quantum computer that will revolutionize the economy, industry, and security. It funds research such as a project at Osaka University to “develop elemental technologies for networking quantum computers with photons, atoms, semiconductors . . . aiming to network small and medium quantum computers” and to “further promote networked quantum computers on a larger scale toward the achievement of universal quantum computation” [44] (p. 10).

Moonshot Goal 7 involves establishing sustainable care systems to overcome major diseases and enable enjoying one’s life with relief from health concerns until age 100. Projects to realize it include one that aims to design technologies that induce sleep and hibernation, and another to develop a method for regenerating lost limits and restoring age-related tissue degeneration.

The general picture that emerges from Japan’s moonshot goals and supporting projects is that the government is funding multiple efforts to introduce next-generation robots, using AI as the main tool for giving robots autonomy, expanding work ranges, and enhancing robots’ ability to function in unstructured spaces. The seven moonshot goals and the supporting projects associated with them are one way to elaborate on the positive values Jobin et al. identified more fully, as shown in Table 3.

Table 3. Positive Values in International AI Ethics Guidance.

| Values | Benefits Gained | Elaboration |
|------------------|---|--|
| Beneficence | Human flourishing and well-being | <ul style="list-style-type: none"> • Moonshot goal 1: Cybernetic avatars • Moonshot goal 2: Ultra-early disease prediction/prevention • Moonshot goal 2: Ultra-early disease prediction/prevention |
| Freedom/Autonomy | Freedom to pursue activities people choose | <ul style="list-style-type: none"> • Moonshot goal 3: Autonomous robots |
| Trust | Positive robot-human relationships | <ul style="list-style-type: none"> • Moonshot goal 3: Robots humans feel comfortable with • <i>Kawai</i> aesthetic |
| Sustainability | Environmental protection and sustainable development | <ul style="list-style-type: none"> • Moonshot goal 4: CO₂ capture, Detoxifying nitrogen, Initiation switches in plastics • Moonshot goal 5: Digitally designed crops, Hyper-customization • Moonshot goal 6: Quantum computing |
| Dignity | Supporting threshold human capabilities | <ul style="list-style-type: none"> • Moonshot goal 7: Overcoming major diseases, restoring age-related loss |
| Solidarity | Considering robots as part of valued relationships and a good society | <ul style="list-style-type: none"> • Moonshot goal 7: Integrating robots in families and natural environments |

Japan's moonshot goals and associated research projects suggest one way to elaborate the positive values uncovered in Jobin et al.'s report. Japan's longstanding tradition of assuming favorable stances toward robots and AI contributes to envisioning a future society that optimizes robots and AI. As Nitto notes, "Many Japanese people who grew up watching robot anime such as . . . Mighty Atom and Doraemon-Gadget Cat from the Future in their childhood tend to define a robot as a partner having a human-like shape and living together with humans" [45] (p. 4). From this vantage point, ethics guidance would be remiss if it did not adequately consider how AI and robots can contribute to solving human problems and creating a better society.

To better incorporate Japanese values into international AI ethics guidelines, we propose combining the precautionary values from Table 1 with the positive and aspirational values from Table 3, resulting in a more balanced and comprehensive AI ethics framework summarized in Table 4.

Combining both Japanese and Western values in international ethics guidance can inch us closer to realizing a more international ethics guidance. Admittedly, a hybrid approach can yield tensions; however, it also promises a more nuanced consideration of practical ethical challenges. To manage these tensions, a hybrid approach ought to use caution and optimism strategically, each serving as a counterweight to the other. For example, the West's precautionary stance serves as a check on Japan's technological exuberance by appealing to worst-case scenarios; Japan's technological optimism serves as a reminder that some risks are worth taking because they lead to better, more flourishing lives.

Table 4. Combining Japanese/Western Values in International Ethics Guidance.

| Values | | Outcomes |
|---------------|------------------|---|
| Precautionary | | Harms Avoided |
| (1) | Transparency | Deception, exploitive and manipulative uses of data |
| (2) | Justice/Fairness | Arbitrary bias and discrimination, Harming marginalized groups |
| (3) | Non-maleficence | Physical and psychological harms |
| (4) | Responsibility | Non-accountability due to diffusion of responsibility |
| (5) | Privacy | Privacy violations, Transgressing boundaries, Misusing persona data |
| Aspirational | | Benefits Gained |
| (6) | Beneficence | Human flourishing and well-being |
| (7) | Freedom/Autonomy | Freedom to pursue activities people choose |
| (8) | Trust | Positive robot-human relationships |
| (9) | Sustainability | Environmental protection and sustainable development |
| (10) | Dignity | Support for threshold human capabilities |
| (11) | Solidarity | Robots incorporated in valued relationships and a good society |

Since we live in an increasingly globalized world, where technologies rapidly diffuse across borders, solutions to ethical challenges from AI and robotics cannot be found solely within one nation state or one system of governance. Global ethical thinking should seek to avoid considering Western ethics “a ‘base’ to be ‘added to’ and thus rendering non-Western values peripheral; instead, a preferred approach is for ethicists, especially Western ethicists, to “actively engage in debates with those from different traditions [than] their own” [46] (p. 315). By more fully elaborating minority values, we can better realize global ethical thinking. ÓhÉigearthaigh et al. point out that thinking more expansively about values can build trust and prevent misunderstanding; they maintain that misunderstanding about AI ethics, not disagreement, is a principal source of distrust between North American and East Asia [47].

5. Conclusions

In conclusion, this paper explored differences between Japanese and Western approaches to robots since the Second World War. It considered how international ethics guidance for AI and robotics negotiates these differences by favoring Western values. We further elaborated Japanese stances and proposed a hybrid approach to ethics guidance that better incorporates them. By bridging differences between Japan and the West, we can better achieve a more ‘international’ ethics for AI and robotics.

Author Contributions: Conceptualization, N.S.J.; writing—original draft preparation, N.S.J. and E.N.; writing—review and editing, N.S.J. and E.N.; project administration, N.S.J. and E.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to Akira Akabayashi for valuable feedback on the paper. This paper was presented at the XXXVII International Congress on Law and Mental Health in July 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Coeckelbergh, M. The Ubuntu Robot: Towards a Relational Conceptual Framework for Intercultural Robotics. *Sci. Eng. Ethics* **2022**, *28*, 16. [CrossRef]
2. Jobin, A.; Ienca, M.; Vayena, E. The Global Landscape of AI Ethics Guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
3. Chattopadhyay, S.; De Vries, R. Bioethical Concerns are Global, Bioethics is Western. *Eubios. J. Asian Int. Bioeth.* **2008**, *18*, 106–109. [PubMed]
4. Bamford, R. Decolonizing Bioethics via African Philosophy. In *Debating African Philosophy*; Hull, G., Ed.; Routledge: New York, NY, USA, 2019; pp. 43–59.
5. Davis, B.W. Opening Up the West: Toward Dialogue with Japanese Philosophy. *J. Jpn. Philos.* **2013**, *1*, 57–83. [CrossRef]
6. Fox, R. The Evolution of American Bioethics. In *Social Science Perspectives in Medical Ethics*; Weisz, G., Ed.; University of Pennsylvania Press: Philadelphia, PA, USA, 1990; pp. 201–217.
7. Fayemi, A.K.; Macaulay-Adeyelu, O.C. Decolonizing Bioethics in Africa. *BEOnline* **2016**, *3*, 68–90. [CrossRef]
8. Čapek, K.R.U.R. *Rossum's Universal Robots*; Wyllie, D., Translator; Feedbooks Publishing: Paris, France, 1921.
9. Oxford University Press. Robot, n.1. In *OED Online*; Oxford University Press: New York, NY, USA, 2021.
10. Science Museum of London. Friend or Foe? Robots in Popular Culture. 2018. Available online: <https://www.sciencemuseum.org.uk/objects-and-stories/friend-or-foe-robots-popular-culture> (accessed on 27 July 2022).
11. Jecker, N.S. My Friend, the Robot; An Argument for E-Friendship. In Proceedings of the Institute of Electrical and Electronic Engineers (IEEE) Conference on Robot and Human Interactive Communications (RO-MAN), Vancouver, BC, Canada, 8–12 August 2021; pp. 692–697. [CrossRef]
12. Elder, A. Robot Friends for Autistic Children. In *Robot Ethics 2.0*; Lin, P., Abney, K., Jenkins, R., Eds.; Oxford University Press: New York, NY, USA, 2017; pp. 113–125.
13. Turkle, S. *Alone Together: Why We Expect More from Technology and Less from Each Other*; Basic Books: New York, NY, USA, 2011.
14. Sparrow, R. The March of the Robot Dogs. *Ethics Inf. Technol.* **2002**, *4*, 305–318. [CrossRef]
15. Heidegger, M. The Question Concerning Technology. In *The Question Concerning Technology and Other Essays*; Lovitt, W., Translator; Harper & Row: San Francisco, CA, USA, 1977.
16. Gates, B.A. Robot in Every Home. *Sci. Am.* **2007**, *296*, 58–65. [CrossRef]
17. Singer, P.W. *Wired for War*; Penguin Books Ltd.: London, UK, 2009.
18. Coeckelbergh, M. The Moral Standing of Machines: Toward a Relational and Non-Cartesian Hermeneutics. *Philos. Technol.* **2014**, *27*, 61–77. [CrossRef]
19. Jecker, N.S. You've Got a Friend in Me: Sociable Robots for Older Adults in an Age of Global Pandemics. *Ethics Inf. Technol.* **2020**, *23* (Suppl. 1), 35–43. [CrossRef]
20. Jecker, N.S. Can We Wrong a Robot? *AI Soc.* **2021**. Ahead of Print. [CrossRef]
21. Meacham, D.; Studley, M. Could a Robot Care. In *Robot Ethics 2.0*; Lin, P., Abney, K., Jenkins, R., Eds.; Oxford University Press: New York, NY, USA, 2017; pp. 97–112.
22. Schodt, F.L. *The AstroBoy Essays*; Stone Bridge Press: Berkeley, CA, USA, 2007.
23. Robertson, J. *Robo Sapiens Japonicus*; University of California Press: Oakland, CA, USA, 2018.
24. Sdshamshel. An Ally of Justice, A Subordinate of Evil, A Symbol of the Past and the Future: 2004's Tetsujin 28. Ogiue Maniax. Anime & Manga Blog. 2010. Available online: <https://ogiuemaniax.com/2010/04/13/an-ally-of-justice-a-subordinate-of-evil-a-symbol-of-the-past-and-the-future-2004s-tetsujin-28/> (accessed on 27 July 2022).
25. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Ageing 2019*; United Nations: Geneva, Switzerland, 2019; p. 12.
26. Chow, L. Will Robots Nannies Save Japan's Economy? In National Public Radio, Planet Money, Public Radio Network, 19 July 2013. Available online: <https://www.npr.org/sections/money/2013/07/19/203372076/will-robot-nannies-save-japans-economy> (accessed on 26 July 2022).
27. Onda, O. Japanese Society 5.0: Going Beyond Industry 4.0. *Bus. Econ. J.* **2019**, *10*, 2–6.
28. Council for Society, Technology and Innovation (n.d.). Society 5.0. Available online: https://www8.cao.go.jp/cstp/english/society5_0/index.html (accessed on 22 March 2022).

29. Strategic Council for AI Technology. *Artificial Intelligence Strategy*; Strategic Council for AI Technology: Tokyo, Japan, 2017. Available online: https://ai-japan.s3-ap-northeast-1.amazonaws.com/7116/0377/5269/Artificial_Intelligence_Technology_StrategyMarch2017.pdf (accessed on 27 July 2022).
30. Kavanagh, C.M.; Jong, J. Is Japan Religious? *J. Study Relig. Nat. Cult.* **2020**, *14*, 152–180. [[CrossRef](#)]
31. Reader, I. Turning to the Gods in Times of Trouble: The Place, Time and Structure of Japanese Religion. In *Reader I, Religion in Contemporary Japan*; Palgrave MacMillan: London, UK, 1991.
32. Gal, D. Perspectives and Approaches in AI Ethics: East Asia. In *The Oxford Handbook of Ethics of Artificial Intelligence*; Pasquale, F., Das, S., Eds.; Oxford University Press: New York, NY, USA, 2020; pp. 607–624.
33. Mori, M. The Buddha in the Robot. Terry, C.S., Translator; Kosei Publishing Co.: Tokyo, Japan, 1981.
34. Kuroda, T.; Dobbins, J.C.; Gay, S. Shinto in the History of Japanese Religion. *J. Jpn. Stud.* **1981**, *7*, 1–21.
35. Kaplan, F. Who's Afraid of the Humanoid? *Int. J. Hum. Robot.* **2004**, *1*, 465–480. [[CrossRef](#)]
36. Mori, M. The Uncanny Valley. *IEEE Robot. Autom. Mag.* **2012**, *19*, 98–100. [[CrossRef](#)]
37. Kasulis, T. Japanese Philosophy. In *Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; Stanford University Press: Stanford, CA, USA, 2019. Available online: <https://plato.stanford.edu/archives/sum2019/entries/japanese-philosophy/> (accessed on 27 July 2022).
38. Jensen, C.B.; Blok, A. Techno-animism in Japan. *Theory Cult. Soc.* **2013**, *30*, 84–115. [[CrossRef](#)]
39. Kovacic, M. The Making of National Robot History in Japan. *Crit. Asian Stud.* **2018**, *5*, 572–590. [[CrossRef](#)]
40. Johnson, C.; Tyson, A. *People Globally Offer Mixed View of the Impact of Artificial Intelligence, Job Automation on Society*; Pew Research Center: Washington, DC, USA, 2020; Available online: <https://www.pewresearch.org/fact-tank/2020/12/15/people-globally-offer-mixed-views-of-the-impact-of-artificial-intelligence-job-automation-on-society/> (accessed on 27 July 2022).
41. United Nations Environment Programme. *Rio Declaration on Environment and Development*; United Nations: Geneva, Switzerland, 1992. Available online: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_CONF.151_26_Vol.I_Declaration.pdf (accessed on 27 July 2022).
42. Resnik, D.B. Is the Precautionary Principle Unscientific? *Stud. Hist. Philos. Biol. Biomed. Sci.* **2003**, *34*, 329–344. [[CrossRef](#)]
43. Hughes, J. How Not to Criticize the Precautionary Principle. *J. Med. Philos.* **2006**, *31*, 447–464. [[CrossRef](#)]
44. Japan Science and Technology Agency (n.d.). Moonshot Goals, 7 Goals. Available online: <https://www.jst.go.jp/moonshot/en/pr/index.html> (accessed on 27 July 2022).
45. Nitto, H.; Taniya, D.; Inagaki, H. Social Acceptance and Impact of Robots and Artificial Intelligence. *Nomura Res. Inst. Pap.* **2017**, *211*, 1–5.
46. Widdows, H. Is Global Ethics Moral Neo-Colonialism? *Bioethics* **2007**, *6*, 305–315. [[CrossRef](#)] [[PubMed](#)]
47. ÓhÉigeartaigh, S.S.; Whittlestone, J.; Liu, Y.; Zeng, Y.; Liu, Z. Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance. *Philos. Technol.* **2020**, *33*, 571–593. [[CrossRef](#)]