Philosophy & Technology (2020) 33:613–638
https://doi.org/10.1007/s13347-020-00421-8

RESEARCH ARTICLE

Harmonizing Artificial Intelligence for Social Good

Nicolas Berberich[1] · Toyoaki Nishida[2] · Shoko Suzuki[3]

Received: 30 September 2019 / Accepted: 4 August 2020 / Published online: 1 September 2020
© The Author(s) 2020

Abstract

To become more broadly applicable, positions on AI ethics require perspectives from non-Western regions and cultures such as China and Japan. In this paper, we propose that the addition of the concept of harmony to the discussion on ethical AI would be highly beneficial due to its centrality in East Asian cultures and its applicability to the challenge of designing AI for social good. We first present a synopsis of different definitions of harmony in multiple contexts, such as music and society, which reveals that the concept is, at its core, about well-balanced relationships and appropriate actions which give rise to order, balance, and aesthetically pleasing phenomena. The mediator for these well-balanced relationships is *Takt* which is an ability to act thoughtfully and sensibly according to the specific situation and to put things into proportion and order. We propose that the central challenge of building harmonizing AI is to make intelligent systems tactful and also to design and use them tactfully. For an AI system to become tactful, it needs to be able to have an advanced sensitivity to the specific contexts which it is in and their social and ethical implications and have the capability of approximately inferring the emotional and cognitive states of people with whom it is interacting.

Keywords Ethics of AI · Harmony · Takt · Human-technology interaction · Technological mediation

✉ Nicolas Berberich
n.berberich@tum.de

Toyoaki Nishida
toyoaki.nishida@gmail.com

Shoko Suzuki
shoko.suzuki.ue@riken.jp; suzuki.shoko.5c@kyoto-u.ac.jp

[1] RIKEN AIP, Technical University Munich, Munich, Germany

[2] University of Fukuchiyama, Fukuchiyama, Kyoto, Japan

[3] RIKEN AIP, Kyoto University Graduate School of Education, Kyoto, Japan

1 Introduction

Artificial intelligence (AI) and its subdiscipline machine learning (ML) have made substantial progress in the last years. Especially techniques for training deep neural networks have been successful enough to make the jump from research laboratories to products and services in a variety of industry sectors. Being used by billions of people on a daily basis, AI technologies already have a major impact on society and considering the many application areas that researchers from academia and industry are working on, from early cancer detection to autonomous cars and robots, this impact will likely grow much larger in the next years and decades. However, technological progress is not deterministic but can and ought to be shaped responsibly and ethically. Carving out what these two terms mean and how they can be achieved for AI technology is the goal and mission of scholars in the applied ethics field called "AI ethics." In the last years, the field has seen a large number of principles of ethical AI being proposed by entities from civil society, industry, and governments. In fact, there have been so many proposed lists of principles that it became hard to keep track and translate them into concrete research directions. In a recent paper, Floridi and Cowls have condensed the proposed principles to a framework of "five core principles for ethical AI": beneficence, non-maleficence, autonomy, justice, and explicability (Floridi and Cowls 2019). The first four are well-known principles from bioethics, whereas the last one, explicability, was introduced as a new principle specific to the domain of AI. However, as Floridi and Cowls mention themselves, the original lists of principles which they used as the starting point of their ethical principal component analysis "emerged either from initiatives with global scope, or from within western liberal democracies" (Floridi and Cowls 2019). For broader applicability, which seems desirable if not outright necessary in our globalized and highly interconnected world, the perspective of other regions and cultures such as east Asia should be incorporated.

This leads us to the main question of this paper:

> Which core ethical principle reflects the Eastern philosophical tradition and can be a valuable asset within a conceptual framework towards ethical development and use of AI technology?

The answer we propose is *harmony*, a concept which originated in music and was later applied to society by Confucius. Since the time of the grand Chinese philosopher, harmony has been a central part of East Asian thought and culture. Due to our own perspective, we focus mainly on the role of harmony in Japanese society and would like to invite Chinese scholars to add their perspectives. However, the *Harmonious Artificial Intelligence Principles*, devised under the lead of Yi Zeng and supported by the Chinese Academy of Science as well as the choice of "HarmonyOS" as the name for Huawei's upcoming operating system, which might well become a sort of national operating system, suggest that harmony still plays an important role in Chinese society. It is not our intention to endorse nor argue against any specific vision of what constitutes a harmonious society. Rather, we want to propose that taking a deep look into the general philosophical concept of harmony and its relation to artificial intelligence is a worthwhile endeavor. Which specific dimensions and interpretations of harmony are the most desirable and technically actionable should be discussed on this general basis and through further research.

To argue for the importance of harmony and its applicability to the challenge of making AI technology ethical, we have structured this paper into two parts. In the first part, we start by taking a closer look into what harmony actually means, where the concept originated and in which contexts it is used nowadays. The concept of harmony is multiform and used in many different areas—from music, mathematics, and art to complex systems and society. Our comparative analysis is mainly focused on the commonalities of how harmony is understood in those different fields, instead of the differences, and will illustrate that harmony is at its core all about relationships between entities within a well-tuned balance according to partly objective, partly aesthetic judgment. We introduce *Takt* as a behavior-guiding perceptional skill which is employed to smooth and thus harmonize social interactions by understanding the particularities of a specific situation as well as the mental state of one's interaction partner.

In the second part, we apply this understanding of harmony to the field of AI. First, we present the Japanese idea of a "Convivial Society of Harmony between Humans and AI" proposed by Yoh'ichi Tohkura and expressed within the Japanese Government's 5th Science and Technology Basic Plan on the "Society 5.0." We then take the common elements of harmony distilled from its different use contexts in the first part—relationships, interactions, and balance—and discuss how they relate to the field of artificial intelligence and what they can offer for designing and using intelligent systems for social good. This analysis will show how harmony provides new and useful perspectives on the ethics of AI which have previously been neglected and gives concrete guidance on meaningful and promising research directions.

Before we begin, we would like to mention that including the concept of harmony in the context of ethics of technology is the merit of Pak-Hang Wong in his paper "Dao, Harmony and Personhood: Towards a Confucian Ethics of Technology" (Wong 2012). Our approach differs from his in that we take a broader view on harmony, including its usage in music, mathematics, and art, but focus on ethics of AI instead of the whole field of ethics of technology and introduce the philosophical concept of Takt as a means of how harmonious interactions can be induced.

The ambiguity in the title of this paper is intended. We have to make intelligent systems more harmonious (we harmonize AI) by tactful design and integration into society, but at the same time, we have to be conscious of the mediating role that AI plays in our sociotechnical society. We should aim to build AI systems that help us to achieve better harmony between ourselves (AI harmonizing us by mediating our interactions).

2 Harmony in Music, Mathematics, and Art

Music The origin of the concept of harmony lies in music. This is reflected by its etymology based on the Greek *harmonikos* which means "skilled in music." Several complementary tones are played together and thus constitute chords that sound harmonious to us. While harmonious music has a strong phenomenological character, it also has an objective character in the form of rules. Tones sound good together when their frequencies form a simple fraction. Their harmony arises not through themselves but through their relationships. An example is the combination of the notes *C* and *G*, which sound harmonious together as their ratio is a simple fraction: G / C = 392 Hz / 262 Hz = 3 / 2. At the same time, as the individual tones create the whole chord through their relationships and differences, they themselves obtain their own identity through these inter-relationships and their situatedness within the whole. In a harmonious system, the absolute position of an entity is dwarfed in importance by its static and dynamic relationships. The mentioning of dynamic relationships brings us to a rhythm. While the concordance of tones in accords constitutes the horizontal dimension of musical harmony, rhythm flows in the vertical direction of musical notation. Again, the essence of rhythm lies not in the elements (the tones) but in their temporal relationships giving rise to a joint repeating pattern.

The element that binds rhythm and concordance together is beat. In this paper, we use its German term *Takt* instead, because it has a broader etymological meaning, including the English concept of *tact*, and allows us to build on its rich tradition in German philosophy.

Takt brings temporal structure to a musical piece and represents its basic unit of time. It is extremely important for harmony as it brings order into music by synchronizing individual voices by virtue of a common temporal structure. In an orchestra with many diverse players, it is a challenge to play their notes simultaneously such that the resulting consonance creates a harmonious result. The solution to this challenge is to be found in the role of the conductor, who marks and thus controls time by waving his or her baton. Furthermore, the conductor chooses the relationships between different voices by tempering some and bringing forth others. His or her role is to weave the contributions of the individual players together into harmony.

As illustrated in Fig. 1, *Takt* plays the role of the shared medium which connects the unique with the general—the single voices with each other and with the overall play and also consonance with the rhythm. It is the magical wand for defining and balancing relationships.

The term *Takt* is derived from the Latin *tactus* which means impact, effect, and sensation (Suzuki 2010). Fittingly, the *Takt* is what we *feel* the strongest when

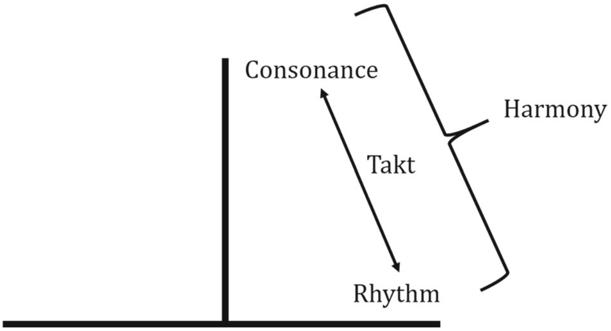


Fig. 1 *Takt* (beat) brings order into music by mediating between consonance and rhythm to produce harmony. The joint sense of Takt allows musicians to play harmonically together

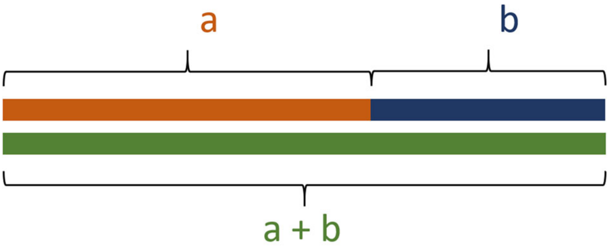
listening to music and the force that brings listeners to tap their toes and nod their heads rhythmically. A related word is *tactile* which refers to the sense of touch. Harmonic music is capable of "touching" us on multiple levels of our being—a feat that harmonious relations and relationships possess in general.

Mathematics According to the influential Russian mathematician Andrey Kolmogorov, mathematics as a scientific discipline originated in ancient Greece in the sixth and fifth centuries BC (Stakhov 2009 p. xix). In these times, mathematics was strongly connected with harmony, without which, according to the Pythagoreans, the Cosmos could not exist (Stakhov 2009 p. xxv). Aristotle later described the position of the Pythagoreans in his *Metaphysics* as the reduction of the universe to harmony and numbers. This worldview of harmonious cosmology was shared by his teacher Plato who considered the regular polyhedrons (platonic solids) as harmonious figures connected with all elements of the universe. Almost 2000 years later, Johannes Kepler developed his platonic solid model of the solar system based on this idea in his *Mysterium Cosmographicum*. Euclid's *Elements* further elaborated on the concept of the harmonic regular polyhedrons and connected them to the *golden ratio* (the division in the extreme and mean ratio). As Fig. 2 illustrates, two parts are in the golden ratio when their relation is equal to the relation between their sum to the larger part. If this proportion is met, the parts and the whole are considered to be in harmony.

Based on his analysis of the role that mathematical harmony played in ancient Greece, Stakhov puts forth the following thesis:

> Euclid's *Elements* was the first attempt to create the *Mathematical Theory of Harmony* which was the main idea of Greek science.
> (Stakhov 2009 p. xxvii)

In contrast to other forms of harmony, this mathematical perspective on harmony focusses on the quantitative side of harmony and thus on numerical proportions. Important in the context of this paper is that just like in music, at its core, harmony is all about relations and about how the parts are related to and situated within the whole. This is further emphasized in a definition of mathematical harmony in the Great Soviet Encyclopedia:



$$\frac{a}{b} = \frac{a+b}{a} = 1.618 \ldots = \varphi$$

Fig. 2 The golden ratio: the ratio between the parts equals the ratio between the whole and a part

> The harmony of an object is a proportionality of the parts and the whole, a merge of the various components of the object to create a uniform organic whole. In harmony, the internal order and the measure of the object had obtained external revealing.
>
> (Stakhov 2009) citing the definition of mathematical harmony in the *Great Soviet Encyclopedia*

Art Similar to many other ideas from ancient Greece, the Italian Renaissance brought new life to the concept of harmony and the golden mean. Luca Paccioli's mathematical treatise on the golden ratio, *Divina Proportione*, was published in 1509 with illustrations by Leonardo da Vinci. This connection inspired the harmonious analysis of Leonardo's artistic masterpieces such as his *Mona Lisa* where ample use of the golden ratio was found. The same was discovered for many other paintings and sculptures created in the Renaissance and earlier in ancient Greece. A particular famous example is the *Doryphorus* by Polyclitus. To study the ideal human body (similar to Leonardo's work on the *Vitruvian Man*), Polyclitus created a sculpture based on three golden ratios (as analyzed by the Russian architect G.D. Grimm in his work *Proportionality in Architecture*). The most important amongst them was the naval partitioning of the body according to the golden ratio. An fMRI study which presented the *Doryphorus* and other Classical and Renaissance sculptures to participants without prior experience in art criticism gave evidence to the hypothesis that there are two processes of perceiving beauty. A process for objective beauty such as exemplified by artistic works based on the golden ratio, which was correlated with neural activity in the insula and a process for perceiving subjective beauty through emotional experiences connected with the activation of the amygdala (Di Dio et al. 2007).

Another occurrence of harmony and the golden ratio in art and design is represented by the Fibonacci spirals. The mathematician Binet showed that the nth Fibonacci number can be expressed in terms of n and the golden ratio: $F_n = \frac{\phi^n - \psi^n}{\phi - \psi}$ with $\psi = 1 - \phi$. Furthermore, Johannes Kepler observed that the ratio of consecutive Fibonacci numbers converges towards the golden ratio ϕ. This forms the basis of a geometric approximation of the golden ratio by successively tiling a plane according to the Fibonacci numbers. Adding circular arcs connecting the opposite corners of the Fibonacci squares results in the Fibonacci spiral (shown in Fig. 3), a geometric figure that has been found in many natural objects and phenomena, from nautilus shells and hurricanes to galaxies.

3 Harmony in Society

3.1 Origin in Confucian Philosophy

The concept of harmony plays a major role in Confucian philosophy. For Confucius, harmony was reflected in ideal relationships, but instead of looking at musical relationships, he focused on social relationships. These relationships could be considered a reflection of the primary relationship between human beings and Heaven and were

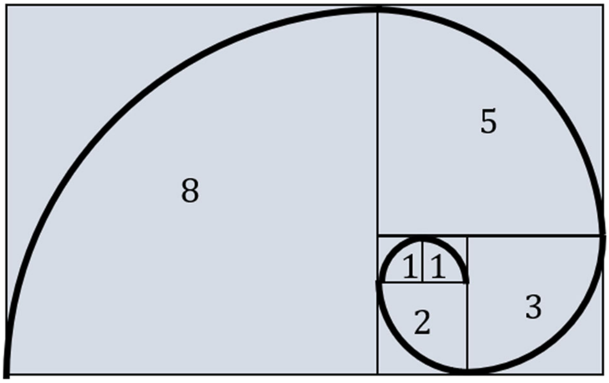


Fig. 3 The Fibonacci spiral: Approximating the harmonious golden ratio

to be found both between individuals (interpersonal) and within individuals (intrapersonal). Especially the family as the social nucleus from which humans learn how to interact with and relate to others appropriately was considered of high importance.

As Pak-Hang Wong notes, "harmony can be conceived as a normative standard of Confucianism" (Wong 2012) because it plays a major role in defining the "good life" and what is required for human flourishing.

Kam-por Yu summarized the notion of harmony in Confucianism through four distinctive features (Yu 2010):

1. Harmony is balancing one thing with another thing
2. Harmony is not complete agreement
3. Harmony is not unprincipled compromise
4. Harmony is the mutual complementation of acceptance and rejection

Balancing is the act of weighting things appropriately and thus creating stable relationships. This notion of Confucian social harmony is in agreement with the abstract definitions of harmony in music, art, and mathematics, which we have presented before. A new perspective is given by the second and third key features of Confucianism, because in English the word *harmony* is often connotated with sameness and avoidance of conflict at any cost. However, in Confucianism, there is a clear distinction between sameness (*Tong*) and harmony. Wong attributes the value that is given to difference to the importance of creative dynamics in Confucianism, which are believed to be essential to human flourishing. Only if there is a diversity of opinions, ways of living, and ideas and they are not killed off in an unprincipled fashion to create a complete agreement, can they be dynamically be combined to complement, support, and enrich each other.

To better explain how this can be achieved, the Confucian classics present the analogy of music which requires, as we have discussed before, similar coordination between its elements. Wong states that in Confucianism this coordination requires the elements "(1) to perform their own roles and functions, (2) to relate to other elements in an appropriate way and (3) not to over-power, or even dominate, other elements."

(Wong 2012). Instead of over-powering, “a genuine harmonious relationship must be backed by reasons” (Wong 2012), similar to humanistic traditions.

More important and interesting than the final state of affairs, i.e., the well-balanced relationships, is the process of creating them. This is why we think that speaking of “harmonizing AI” instead of “harmonious AI” is more appropriate and constructive. As Wong further elaborates in his paper, relating to Yu’s fourth key feature of Confucianism, the way to achieve harmony is through a process of continuous negotiation and mutual adjustment. This reminds of discourse ethics and its focus on communicative processes as described by Juergen Habermas and Karl-Otto Apel.

The discussion of harmony in music, mathematics, art, and society has shown multiple commonalities such as a strong emphasis on balanced relationships, concerted interactions, a connection between rule-based and subjective aesthetics, and the interplay between the general and the particular. However, there are differences as well. Musical and societal harmony are dynamic properties which unfold through time and are co-constructed by different agents, while the concept of harmony in mathematics and Renaissance art is based on static proportions. AI is distinct as a technology in its ability to dynamically interact with humans rather than being a static artifact. Therefore, we will focus on the social perspective on harmony (which, as introduced before, was influenced by musical concepts) for the rest of the paper.

In addition to harmony, several other terms from music are regularly applied to describe societal phenomenons. Consonance between people refers to an agreement or at least a compatibility between their opinions and actions. Another concept that is originally known from music and has been applied to social interactions is the idea *Takt*, which we will discuss in the next section.

3.2 Takt in Social Situations

The first metaphorical use of tact, not as the tactile sensation of physical contact, but as the perception of social contact and appropriate behavior directed towards the avoidance of offense or embarrassment, appeared in a letter written by Voltaire in 1769 (Heyd 1995). However, as a philosophical concept, it is a child of the Romantic era, during the end of the eighteenth and the first half of the nineteenth century in Europe, with its emphasis on the particularity of situations and the importance of emotions and aesthetics. It has been especially central in the work of Johann Friedrich Herbart (1776-1841), who followed Kant as professor of philosophy at Königsberg University and established pedagogy as an academic discipline. Herbart’s *pädagogischer Takt* is the faculty of judgment (*Urteilskraft*) with which teachers mediate the *harmonic balance* between theory (the curriculum) and practice (the particular instructional situation) (Suzuki 2008; 2010). As a tutor for the children of a wealthy household in Switzerland, Herbart had learned through his own experience how important the skill of pacing the educational interactions with one’s students is. He learned to perceive when it was the time to talk and when it was the time to let his students talk, and most importantly, when to give the student an important idea which they would be in the mood and state of mind to further develop (Suzuki 2008). In other words, Takt is the ability to perceive the cognitive and emotional state of the students and to judge based on this understanding of how the general lesson plan can be put into practice.

When there is a conflict between students, teachers need to apply a high level of *Takt*. They usually do not have complete information about what happened between the students or which underlying issues might fuel the conflict, but still are challenged with the task of resolving the disagreement harmoniously. In mediation, it is often useful to ask the conflicting parties to recount the situation from their own perspective and describe to the other party how it made them feel. The goal is thus to promote dialogue and empathy. The tactfully mediating teacher has to show both of them how this can be done by him or herself empathizing with their position and by helping them articulate their thoughts and emotions while sustaining a neutral position and balancing the attention giving to each of them.

Another area of social life where tactful behavior for harmonious interaction is of high importance is the interaction between patients and doctors. A situation of high moral salience is when the doctor tells the patient about the result of their diagnosis. Especially when the diagnosis is negative, it is important to care for their emotional states and if necessary to comfort them and give explanations of the diagnosis. This needs to be considered when building medical AI systems.

Herbart's pedagogical philosophy has been connected to the work of Kitaro Nishida (1870–1945), the progenitor of the Kyoto School of Philosophy and one of Japan's premier philosophers (Suzuki 2012). Nishida was both directly and indirectly influenced by Herbart's ideas, yet at the same time strongly embedded in traditional Eastern thought by being based on the idea of the human as a dynamic being which is constantly interacting and in contact with its environment.

Due to its etymology, Takt has been described as a "sense organ for touching the outside world" (Suzuki 2019). While the term "sense" reflects its heritage originating in the age of sensitivity, we believe that "perception" serves as a more suitable conceptual framing of Takt. In cognitive science and robotics, there is a difference between sensing and perception. A sensor (in neuroscience: receptor) measures external physical properties such as temperature or force and transduces them into electrical signals which are usually then discretized (either through spiking neurons or through analogue-digital converters). Thus, it is a unidirectional process. Perception on the other hand involves a cognitive component through which attention, expectations, understanding of the context, and prior knowledge guide and even construct what we perceive. Thus, perception goes beyond the sense-think-act loop of classical AI. A tactful person is able to integrate "all five senses into a unified whole" to construct a good understanding of the current situation and the other people within it (Suzuki 2019). Because this perceptional integration process is agent-specific and directed towards the most appropriate physical response to the given situation, it can be understood as a corporeal intelligence.

As we will discuss in the later section on pro-social behavior and AI, embodied AI such as cognitive robots or avatars lend themselves better to tactful behavior than disembodied applications of AI such as chat-bots or pattern classification systems. Without the physical or virtual embodiment, they lack the contact surface (as a side mention, "contact" shares the same etymology as Takt) for rich and considerate interactions. This interactional contact surface is bidirectional, which is why Takt does not only have a perceptive component, but also an active one. Someone does not have takt if they are able to perceive that a certain action would offend someone else, but still choose to perform the action. Thus, takt necessarily includes a second, active component in addition to its perceptive one. Therefore, Takt is a practical intelligence which guides behavior to be more tactful through attentive perception to others and the current circumstances and thus creates harmonious situations and relationships. The dual nature of Takt as a perceptual skill as well as a behavioral skill is of Western origin (as illustrated in Descartes' dualism between mind and body), whereas the Japanese cultural perspective allows it to be seen as a singular concept, similar to their word *mi* which simultaneously connotes the physical body, one's state of consciousness and dynamic social relations (Suzuki 2019). One's lived experiences are engraved in one's body as tacit knowledge similar to body memory. One might argue that this uniquely Japanese perspective of corporeal intelligence fuels the Japanese concentration on embodiment research in AI, especially through neuro-inspired robotics. This difference to the Western mind-focused AI, where even non-interacting data analysis methods are called artificial intelligence, is not yet well reflected in AI ethics.

In more recent philosophical literature, Takt has been described as a behavioral virtue (Heyd 1995; Naukkarinen 2014) which similar to other virtues can only be cultivated through practice under real conditions and through studying the performance of experts. Takt is not a character virtue like the virtues discussed by Aristotle in his *Nicomachean Ethics*. A person who is tactful in one particular situation can be tactless in a different situation, for example, when traveling to a foreign country with an unfamiliar culture. As Ossi Naukkarinen points out, Takt is a relational concept (Naukkarinen 2014). We do not ascribe the property of being tactful to a certain behavior because of its content or appearance alone, but because of its relationship to the particular context in which it is performed. Given an appropriate context, almost any behavior can be tactful, and given an inappropriate context, almost any behavior can be tactless. As David Heyd puts it "[A] tactless saying may be wrong and offensive not due to its content, but rather due to the circumstances of its utterance; not because its motive or intention, but rather because of lack of sensitivity to its impact on others." (Heyd 1995).

Takt has been connected to what Kant called judgment, but as Heyd rightfully pointed out, it is much closer to Aristotle's virtue of *phronesis* (practical wisdom) (Heyd 1995). Similar to *phronesis*, Takt allows to act appropriately in particular situations based on one's selection of the most salient and essential features. Furthermore, analogous to virtues, Takt can be cultivated through experience and is not guided by universal rules. Interestingly, actively tactful behavior is not morally obligatory and not universally expected even though we value it highly. Often, it involves deflecting attention or even silencing others before they can say something that is hurtful or offensive to someone else. Because it is non-obligatory and not rule-based and depends on the specific situation, tactful behavior cannot be theoretically planned in advance but needs the ability to improvise and involves a creative element. Heyd gives an example of Queen Victoria who observed an ill-mannered guest at her dinner table ignorantly drinking from his finger bowl. Without much hesitation, she imitated his behavior by drinking from her own finger bowl and thus saved him from

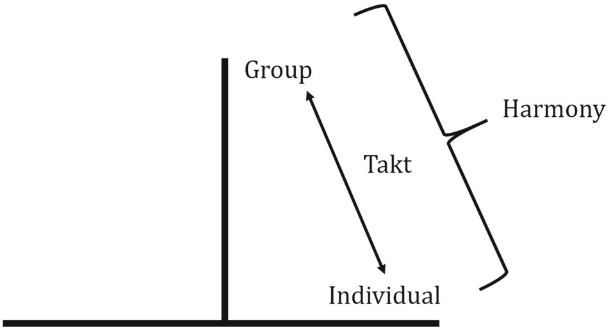


Fig. 4 Harmony in society: The well-balanced relationship between the individual and the group. The sense of Takt allows an individual to sense the general atmosphere of a discussion and to harmoniously join in, while at the same time grasp the particularity of each individual's mental constitution in order to not cause offense or embarrassment

embarrassment. This example, which shows an artistic act of balancing between morals and mores, illustrates the poietic element of Takt.

To borrow Floridi's terminology from Floridi and Sanders (2005), Takt is *ecopoietic* as it does not focus on building one's own character (*egopoietic*) as discussed in virtue ethics, but aims at proactively shaping one's surrounding and especially the behavioral interactions. This observation of Takt and harmony being similar to virtue ethics, but still different, fits well with Heyd's classification of Takt as a behavior instead of a character virtue. Since it specifically deals with harmonizing the interactions with other humans, Takt can be described as *sociopoietic*. As Takt often deals with appropriate behavior when different cultures make contact, it has the potential to scale up from simple personal interactions to societal interactions. For example, tactful recommender systems might recommend certain topics to some cultural groups while not showing them to other cultural groups to avoid offense.

In the German tradition of differentiating between the human sciences (*Geisteswissenschaften*), which are focused on human particularities, and the natural sciences (*Naturwissenschaften*), which concern themselves with natural generalities, Helmholtz, Dilthey, and Gadamer consider Takt to be a unique feature of the human sciences (Heyd 1995). AI understood as "the science and engineering of making intelligent machines" [1] bridges both Snowian cultures under the premise that engineering is understood as designing sociotechnical systems. Thus, the discipline of AI needs to mediate between scientific generalities and human particularities.

In the last sections, we have seen that harmony emerges out of the appropriate balancing between opposites such as consonance and rhythm (Fig. 1), the whole and the parts (Fig. 2), and the group and the individual (Fig. 4). It has been shown that the concept and skill of *Takt* take the role of artful balancing and mediating:

[1] John McCarthy in "What is Artificial Intelligence?" (revised version from 2007) http://www-formal.stanford.edu/jmc/whatisai.pdf

> *Takt* as medium of the identical and the different, the individual and the collective, the cognitive and the emotional, the reasonable and the sensual, the mental and the corporal. (Suzuki 2014) (translated from the German original)

3.3 From Japanese Wa「和」to Reiwa「令和」

In an interview with the New York Times, the Nobel laureate Hideki Shirakawa stated that fundamentally, Japanese culture is based on rice farming. To efficiently plant rice, it is required that teams of people walk from row to row at the same speed (French 2001), dynamically bringing their own actions and the perceived actions of others into proportional balance. In other words, Japanese culture has a strongly ingrained sense of *Takt* and thus of harmony.

The Japanese term for harmony is *Wa* which constitutes the country's most important value as well as being its oldest recorded name. Already the first constitution created in Japan, Prince Shotoku's Seventeen Article Constitution from 604, put emphasize on *Wa* in its first article:

> Harmony [Wa] should be valued and quarrels should be avoided. Everyone has his biases, and few men are far-sighted. Therefore, some disobey their lords and fathers and keep up feuds with their neighbors. But when the superiors are in harmony with each other and the inferiors are friendly, then affairs are discussed quietly and the right view of matters prevails.
> English translation from Aston (1896)

According to Japanese philosopher Takeshi Umehara, this first article has had more influence on the Japanese people than any other (Kramer and Ikeda 1997). Historically, the Japanese ethical concept of *Wa* was derived from the Confucian ideal of harmony when many elements of Chinese culture where introduced to Japan in the fifth century CE. Through contact with Shinto and other Japanese cultural elements, such as a special focus on politeness, it gained its distinct Japanese style. *Wa* is about reliance and trust between people and about making decisions by consensus. It means people working "politely and keeping a good relationship in a group, with full appreciation of the uniqueness of all members to reach the goal: goodness, peace and growth of all members involved" (Konishi et al. 2009).

A prerequisite of politeness and peace in the sense of *Wa* is to know which kind of behavior is appropriate in each given situation and how one's actions might affect the other group members. In other words, the ability of *tactful* behavior is a necessary ability of group members to achieve *Wa*. As Konishi et al. have observed in a study with Japanese nurses (Konishi et al. 2009), the shared wish for harmony bears the risk of confusing harmony with conformity and thus not only failing Confucius' teaching but creating unfortunate situations. The nurses they interviewed were often in disagreement with physicians and institutional rules, but would not voice their objections or alternative ideas due to the difference in rank or seniority. Tactful behavior in the sense of genuine harmony requires the skill to state one's position clearly, but politely and calmly, while expressing respect to the other party. Konishi et al. give examples of Japanese nurses practicing this skill when discussing with physicians about the course of action they believe to be best for the patient.

Perhaps the best explanation of the meaning of *Wa* for non-Japanese people has been given by the linguist Anna Wierzbicka in her work *Japanese Key Words and Core Cultural Values* (Wierzbicka 1991). She broke the concept down into a series of phrases consisting of simple, precise, and culture-independent terms:

Wa
(a) these people think something like this:
(b) we are not one thing
(c) we want to be one thing
(d) we all want the same
(e) we want to do something because of this
(f) we don't want this:
(g) one of us says: "I want this"
(h) another one says: "I don't want this"
(i) we don't want to say:
(j) "one of us did something good,
(k) another one did something bad"
(l) they all feel something good because of this
(m) they can do something good because of this
(n) people think this is good

Analysis of *Wa* in culture-independent terms (Wierzbicka 1991)

Rows (c) and (d) represent the wish for the unity of being and goals, while (g) and (h) the desire for agreement. From an ethical perspective, rows (l) to (n) are of special importance, because they illustrate that *Wa* is both good for the group members that share it, but also a tool for performing good deeds for society.

May 1, 2019, was a historic day for Japan. It marked the day of the beginning of a new imperial era after emperor Akihito had abdicated his reign the day before and his son Naruhito ascended the Chrysanthemum Throne as the new Emperor of Japan. The name for Emperor Naruhito's reign was chosen by the Japanese cabinet from a list of candidate names produced by an expert committee of nine highly regarded members of Japanese society. The name chosen—Reiwa—can be interpreted (but not directly translated) as "beautiful harmony" in English according to the Japanese Foreign Ministry. The chosen name of the era has an important meaning as it summarizes the vision of the country for the next decades and further emphasizes the extraordinary importance that the value of *Wa* holds in Japan's culture.

4 Harmony in AI-Enhanced Sociotechnical Systems

4.1 The Convivial Society of Harmony Between Humans and AI

The new imperial era of *Reiwa* in Japan coincides with the push towards the establishment of a *Society 5.0* which, according to the Japanese Government, will be "a human-centered society that balances economic advancement with the resolution of social problems that highly integrates cyberspace and physical space." (Government of Japan 2016). It is said to follow the hunting and gathering society (Society 1.0), the agricultural society (Society 2.0), the industrial society (Society 3.0), and our current information society (Society 4.0) (see Fig. 5). The technological backbone of this "super smart society" is proposed to be the combination of the internet of things with artificial intelligence and robotics. They should be deployed towards meeting human needs and supporting their activities in a variety of areas, from manufacturing (similar to Germany's Industry 4.0) to transportation, healthcare, sales, and public services.

This idea of subsequent stages of human societies is also present in Yoh'ichi Tohkura's grand conjecture about progress towards Society 5.0 as a convivial society in which humans and intelligent non-human information systems would live in harmony. He suggested that at the core of this sequence of societal types was a shift in the role that information played in human societies. While in early hunter-gatherer society information was used for finding food and staying away from danger. His vision laid the groundwork of a large research program called *Human-Harmonized Information Technology for a Convivial Society*, funded by the Japanese Science and Technology Agency (JST), which started in 2009 under Tohkura's and later Toyoaki Nishida's supervision.

The goal of the research program was to work towards the creation of "basic technology that enables an information environment that is in harmony with people" (Nishida 2016 p. 2). Humans should be in harmony with the information environment (infosphere, see Floridi 2013, p. 8ff) surrounding them, just like Confucius promoted harmony with nature (the biosphere).

As we have discussed before, the combination of, and balance between, the individual and the general is central to the concept of harmony. Aligned with this position, the harmonious convivial society is described as one where both human potential and social potential can actualize and flourish:

> [P]eople will need to find a nontraditional style of self-actualization and society will aspire to a new principle of endorsing harmony. Human potential is the power of an individual that enables her or him to actively sustain an endeavor to achieve a goal in maintaining a social relationship with other people. It

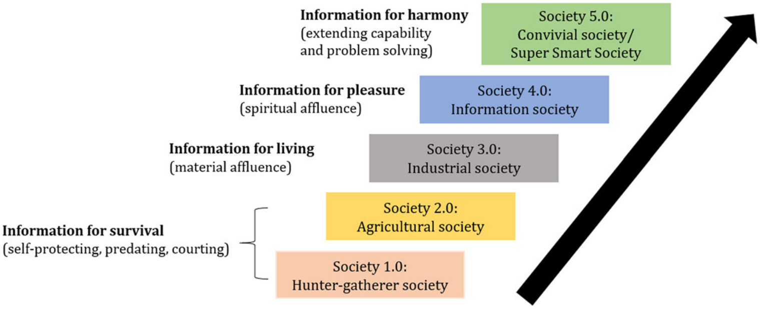


Fig. 5 The stages of society towards the convivial Society 5.0 (adapted from Nishida (2016)). The demarcation between the different states is constituted through the changing role of information in human societies

> involves vision, activity, sustainability, empathy, ethics, humor, and aesthetic sense. Social potential is the power that a society of people possesses as a whole. It encompasses generosity, supportiveness, conviviality, diversity, connectedness, and innovativeness. We believe that human and social potentials complement each other to enable conviviality. (Nishida 2016, p. xii)

The results of the research program are summarized in Nishida (2016) and Nishida (2017). Interestingly, *Takt* played an important role in multiple of the individual research projects, such as in Hiroshi Ishiguro's minimal design approach towards transmitting human presence (*sonyaikan*) through androids (Ishiguro 2016). His research gave evidence that touch and its integration with other modes are integral for enhancing *sonyaikan*. AI systems do not need to be autonomous but can also be semi-autonomous and teleoperated. Ishiguro et al. observed that when introducing their teleoperated robot into an elementary school setting, the distant classmate operating the robot became more integrated into the group's activity in contrast to non-embodied communication tools. As we will discuss in the next section, the teleoperated robotic system mediates the interaction between the school children and is ethically relevant to the extent that it promotes harmony. Similar telepresence systems might in the future be used by bedridden patients and elderly to participate in social groups and contribute towards their harmony. As Matthew Gladded discusses in his recent paper on *Who will Be The Members Of Society 5.0?* these cyber-physical systems might give rise to diverse types of human and non-human members of Society 5.0 (Gladden 2019), reminding of actor-network theory.

4.2 Relationships, Interactions, and AI

Our synopsis of the different use contexts of harmony in the first part of this paper has illustrated that proportional well-balanced relationships are the most important feature of harmony. *Takt* is the ability to act on these relationships and put them into order and proportion. This resembles the central idea of *mediation theory* in the philosophy of technology which has been developed by the Dutch philosopher Peter-Paul Verbeek based on post-phenomenology and builds upon Don Ihde's analysis of human-technology relations (Verbeek 2005; 2011; 2015):

> [The notion of technological mediation] indicates the ways in which technologies inevitably and often implicitly help to shape human actions and perceptions, by establishing relations between users and their environment.
> (Verbeek 2009, p. 66)

AI is a special technological field as its different offspring addresses and mediates more types of relationships than probably any other technological discipline (see Fig. 6). Deep learning–based computer vision applied to satellite images allows us to see the earth from new perspectives, while social media platforms apply machine learning to mediate the social interactions of their users, e.g., by automatically detecting hate speech. AI techniques applied to health and fitness data create new lenses under which we observe, interpret, and remodel ourselves and lastly, conversational AI systems engage in direct interactions with human users. Often, the AI-as-agent (or

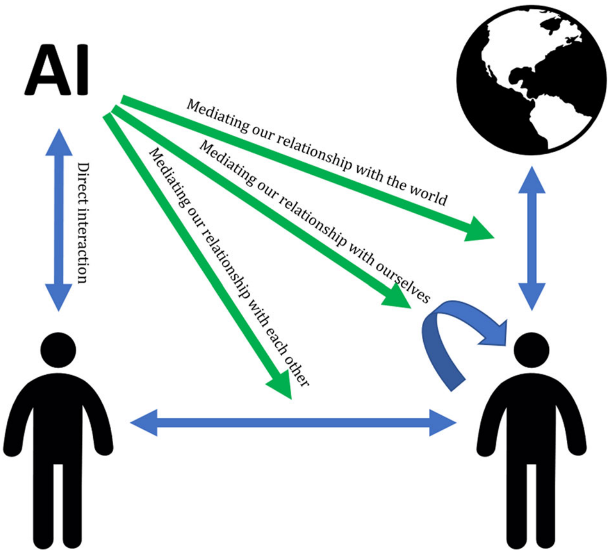

AI
Direct interaction
Mediating our relationship with the world
Mediating our relationship with ourselves
Mediating our relationship with each other

Fig. 6 AI is can both interact with us directly in an agent-like manner, but is also mediating many different relationships such as the relationship between different people, between humans and the natural environment and even how we see and treat ourselves

AI-as-subject) perspective is narrowed down to these "alterity relations (Ihde 1990)" in which humans interact directly with technology and vice versa and is thus solely discussed in terms of machine morality, while the ethical implications of AI-as-a-tool (or AI-as-object) are discussed similar to other technologies. However, as we will show in the following sections, agent-like AI systems such as conversational agents or assistive robots can also mediate other relationships, especially between humans.

First engineering research on this perspective of "intelligent machines as social catalysts" (Rahwan et al. 2020) has already been conducted. One recent example is a study by Margaret Traeger et al. which gives evidence for the hypothesis that a social robot can mediate human-to-human interactions. They found that in cases where a robot was making vulnerable statements within a group consisting of two humans and one robot, the humans tended to converse significantly more with each other and perceived their group more positively (Traeger et al. 2020). In a case study with the therapy robot seal Paro at an elderly care house, Wada and Shibata observed that the robot mediated more communication between the elderly (Wada and Shibata 2007). An even stronger example of research on Takt in the context of harmonizing human relationships through mediating AI is given by Malte Jung et al. in a study on using robots to moderate team conflict (Jung et al. 2015). The robot was programmed to intervene during a group's performance on a problem-solving task by trying to repair task-directed and personal attacks. Unlike a tactful person, who is able to deflect attention from an offensive action the researchers observed that the robot's attempt at moderating team conflict increased the participants' awareness of conflict. While this gives evidence for the mediation hypothesis, it demonstrates that Takt is an artful skill which needs to be applied with delicate sensitivity and dexterity, which current robots and AI systems do not yet possess. The purpose of designing AI systems that exhibit the skill of Takt is not only for human-AI interactions, but also to create technology with a harmonizing mediation effect on human-human interactions.

4.3 Takt, Balance, and AI

We have shown that Takt is the prime measure by which harmony is achieved. Thus, a prime challenge of building harmonizing AI is to make intelligent systems tactful in its interactions with humans. Smart assistants, both virtual and physical, need to know when to engage in interaction with humans and even more important when to disengage and remain silent. An example is a research on an attentive quiz agent that "observes" human group interaction and avoids making an utterance when participants are actively engaged in discussion (Huang and Nishida 2013).

Research on the topic of human-robot coordination dynamics has given evidence that robots affect intentional group coordination and synchrony (Iqbal and Riek 2017). Iqbal and Riek employed several anticipation algorithms with which the robots were supposed to adapt to the Takt in human movement. Their results showed that the addition of one or multiple robots to the human-only group significantly reduced the group coordination, which might pose a challenge to achieving *Wa* in Society 5.0.

We believe that the field of conversational informatics (see Nishida 2008, 2014) can play a central role in teaching intelligent systems about tactful behavior in conversational interaction with humans. Conversational informatics studies the verbal and non-verbal communications of conversation participants with the help of data-intensive recording, data analysis methods, and artificial intelligence. Besides the scientific goal of contributing to a better understanding of how humans share thoughts and feelings through social signals, conversational informatics aims at designing (embodied) conversational agents that can fluently converse with individuals or groups of humans. New human sensing tools based on deep learning such as neural network-based pose tracking, activity, and speech recognition and inference of emotional correlates in speech prosody and facial expressions allow intelligent systems to get a better situational awareness, both about the context and the humans' inferred emotional state, which could be used to design tactful conversational AI. However, these technologies are very intimate and thus require careful deliberation by the engineers about whether they are appropriate for the intended use context or if they infringe on privacy.

Tactful recommender systems should also take into consideration what is socially appropriate. Behavioral measures correlated with adipositas should perhaps, for example, not be used for recommending diet products. And, to use an extreme example, if a woman had a miscarriage, she should not keep getting bombarded by ads of baby products. If the complete situational awareness is not possible for the system, e.g., due to missing data or privacy restrictions, it should at least have a feedback option for the human to stop its tactless actions.

Putting AI into the *Takt* of humans means that the human sets the pace and acts in a humanistic sense as the main regulatory force which secures the balance. The human should remain the conductor of the harmonic orchestra of human-AI interactions by wielding the baton which sets and communicates the Takt. This is not a new idea, but one that needs to be continuously re-emphasized. Mike Cooley, who coined the term *human-centered systems* illustrated the perils when "you are being paced by the machine, and the pace at which you work is becoming more and more visible" (Cooley 1987, p. 38ff). He contrasted the ruthlessly precise pacing of work in the Taylorian fashion where even trips to the lavatory needed to be within time allowances of two-decimal accuracy (1.62 minutes) to natural rhythms of work based on the rhythm of seasons, daylight, and animals physiology. In manufacturing, the most important measure of time is the interval between the start of production of one unit and the start of the next unit, which is fittingly called *Takt time* (in Japanese *takutotaimu*) based on the German *Taktzeit*.

This position of putting AI in the *Takt* of humans is the exact opposite of calls to enhance our brains to "keep pace" with the rapid developments in AI in order for humans not to be left in the dust. Such narratives stage a race between humans and technology without a clear finish line.

The value of exhibiting moderation, restraint, and temperance while living in harmony with nature as described by Confucius appears very timely with regards to the ongoing climate crisis. Harmony-promoting AI should support us in strengthening our relationships with the world and bring it back into balance. Recent research has highlighted the enormous energy consumption necessary to train state-of-the-art natural language processing models (Strubell et al. 2019) and the potentially much higher energy cost that the simulation of true human intelligence—the goal of the early AI pioneers—might require in the future. Meanwhile, other AI scientists argue that AI is not just part of the problem of the ongoing climate change, but could be a powerful tool towards its solution, e.g., by optimizing infrastructure and vehicle efficiency (Rolnick et al. 2019). However, as the authors emphasize, technology alone is not enough and machine learning is not a silver bullet. Instead, a collaboration with domain experts and society will be required—a harmonious act of balancing expertise, experience, tools, expectations, costs, and benefits.

4.4 Pro-social Behaviors and AI

When AI researchers and ethicists discuss AI for social good, they tend to focus on its use in application areas which generally have a direct positive impact on society such as education and healthcare. However, AI can also be socially good on the individual-level as demonstrated by extensive research in the field of socially assistive robotics. Sociologists and psychologists describe socially good behavior between individual people as *pro-social behavior* which includes acts like helping, sharing, and co-operating.

AI systems which encourage pro-social behaviors between people or even perform pro-social behaviors themselves can indirectly lead to socially good outcomes. However, in contrast to the systems that directly aim at the good purposes of the societal systems such as education and healthcare systems, these indirect effects strongly depend on the persons involved. An AI system helping people with unethical or ignorant goals can also produce social ill. One major challenge to avoid bad outcomes is to program the system such that it always considers the effects which it's own pro-social behavior or the pro-social behavior it is mediating have on other people, including people outside the direct social circle of the human interaction partner. This pro-social behavior mediation might, for example, be through tools that make it easier to help others or to collaborate, e.g., the application of machine learning for efficient car-sharing networks.

Tactful behavior can be seen as one specific pro-social behavior which is performed on the individual-level but can give rise to the emergent property of harmony at the group-level or societal-level (with the specificity that it can also include non-actions such as restraint). Tactful AI systems as non-human actors on the level of individuals might contribute to emergent harmony, similar to how the "Swiss Robots" from a seminal experiment by the ethologist Rene te Boekhorst and the engineer Marinus Maris followed simple rules (similar to Braitenberg's vehicles) which resulted in a complex collective behavior which could be interpreted as "cleaning" (Maris and Boeckhorst 1996). If the simple individual behaviors of AI systems are implemented sufficiently well by humans, it might be conceivable that they could contribute to societal harmony even without having any understanding of harmony or Takt by themselves. The Takt that the systems would express would therefore be ultimately based on human sensitivity. Due to the context-sensitivity of Takt, simple rules will not suffice for tactful AI systems. Instead, they will need to be able to adapt, learn, and follow human cues.

In this spirit of human-in-the-loop, harmonizing AI systems need to not only be human-adaptive, i.e., adapting to human behavior and preferences, but also human-adaptable, i.e., giving users the opportunity to directly tell it about their preferences and what they do not like. This could, for example, be recommender systems which are not solely trained on our clicking behavior, but can also be shaped directly through commands.

An important question when setting the goal of a harmonious society is "Who perceives the harmony?" Usually, engineers define a goal metric and then translate it into a cost function (or value function) that can be minimized (or respectively maximized). However, for the case of harmony, this would only be possible if it was completely measurable objectively. If this was possible, one might imagine the development of autonomous harmony-maximizers that perceive and evaluate harmony without human interference. We believe that this image is not desirable and most likely not possible. It is questionable whether the ability to sense harmony can in principle be transferred to machines beyond simplified features such as synchronous group dynamics. As illustrated in Section 3.3 with the example of Japanese *Wa*, harmony is a joint social experience and should therefore be evaluated collectively.

With these challenges for tactful AI described above in mind, we believe that further and strengthened research in the fields of social cognition and collective intelligence with focus on joint actions, group dynamics, and pro-social behavior in relation to intelligent systems such as robots or virtual agents would be highly desirable towards the goal of harmonizing AI.

5 Tactful Harmony as an Ethical Core Principle for AI

To support our recommendation that harmony should be considered as an additional core principle of AI ethics, we argue for three propositions: (1) It affords a valuable perspective, especially in the context of AI systems. (2) It is not already subsumed by one of the other core principles. (3) It does not itself subsume the other principles. Given the premise that valuable perspectives should be included in the core principles, we conclude that harmony and its herald Takt should be included as an additional core principle of AI ethics.

1) It affords a valuable perspective, especially in the context of AI systems. In medicine, a principled approach to reflecting the ethical dimensions of a decision to prescribe a certain therapy would ask the following questions:

- Beneficience: “Will it bring about good, e.g., curing the patient’s illness?”
- Non-Maleficience: “Will it avoid harm, e.g., unnecessary or disproportional side-effects?”
- Justice: “Are we treating this patient impartially and fairly with respect to other patients, i.e., through a Rawlsian veil of ignorance?”
- Autonomy: “Does the patient support this action out of her own will, e.g., through informed consent?”

If, for example, a medical decision support system based on machine learning is supposed to help with the above therapeutic decision, then the additional questions of “How does it work?” and “Who is responsible?” need to be addressed.

- Explicability: “How does it work?” (intelligibility) and “Who is responsible?” (accountability)

Floridi and Cowls subsume these questions under the principle of *explicability*, which they add as a fifth principle to the four original ones from bioethics, thus creating a framework of five principles for AI ethics (Floridi and Cowls 2019).

Let us consider the example of the hard task of telling a patient and their family that the tested treatment has not worked and her cancer is terminal. A system based on machine learning might be able to classify the negative response to the treatment with high accuracy and might utilize natural language processing to inform the patient. However, one might with good reason argue that such a system would be unethical because its utilization is highly tactless. Since Takt is highly situational and relational, it makes a difference *who* says something. In cases like these, human tasks should not be replaced by robots or AI systems—not because of functional reasons, but because of Takt. Takt is thus linked to restraint, not saying or doing something out of empathetic consideration for others. It is the antidote against overuse. If in some situations, AI systems are used to convey hard news to humans, they should do it in a tactful manner. Again, situational specificity plays a role as well as one’s own role in the situation. An AI system or robot trying to express empathy through anthropomorphically mimicking compassion might cause offense by being perceived as deceptive and thus tactless. Hans-Georg Gadamer observed that tactlessness often involves the invasion of privacy, for example, by asking strangers personal questions or by talking about someone's personal information in the presence of someone else with whom they do not want to share their information. Again, the ethical essence lies not in the act itself, but in the social relationships and their situational context in which it takes place.

When looking at AI ethics from an intercultural perspective, it becomes obvious that the tool-centric paradigm which is dominant in the West is not without alternative. In a recent paper, Gal et al. compare the approaches and attitudes towards AI ethics in South Korea, China, and Japan, from the perspectives of policy, academic thought, and popular culture (Gal 2019). They conclude that while in South Korea AI systems and robots are seen as mere tools, in Japan a clear societal vision for harmonious co-existence with these technologies is existent (e.g., the Society 5.0 vision and Sony's AI Ethics Guidelines [2]). China lies between those two positions, leaning towards tools in terms of policy, while Gal et al. describe their academic thought and popular culture as partnership-oriented. Even in the tool-centric perspective, harmony plays an important role in China. A nationally wide taught graduate engineering ethics textbook highlights harmony as one of four unique Chinese characteristics in comparison with Western engineering ethics guidelines (Gal 2019). In the partnership-oriented academic discussions, harmony plays an even greater role as shown by the *Harmonious Artificial Intelligence Principles*, devised under the lead of Yi Zeng. In contrast to most other lists of principles, it puts a strong emphasis on the interaction between humans and AI agents, including empathetic connections. Its main difference in perspective to the Western focus on trustworthy AI[3] is given by the following sentence "AI ethics is not only about how robots and machines provide better and more trustworthy services to and interactions with humanity, but should really about how to construct the harmony Human-AI society."[4]

It is not the goal of this paper to recommend the AI-as-a-partner perspective over the AI-as-tool perspective or vice versa. Instead, we intend to point out that if a unified framework of principles is supposed to afford a structured ethical reflection on AI systems, then it would gain much by incorporating a perspective that discusses the system along the dimensions of tool, mediator, or interaction partner. The discussion about the spectrum between tool and interaction partner is specific to AI, no other technology exhibits a similar level of agency and thus the capacity for dynamic bidirectional interactions with humans. This technological particularity should be reflected in the ethical reflections on the scientific field and its sociotechnical developments. The current framework of five principles furthermore does not reflect these presented cultural differences.

Another valuable contribution which tactful harmony would make to the framework is that it provides a more suitable context to discuss the value of privacy. Floridi et al. have placed privacy as a sub-principle of non-maleficence. However, the

[2] "[...] promoting the utilization of AI in a manner harmonized with society [...]" https://www.sony.net/SonyInfo/csr_report/humanrights/hkrfmg0000007rtj-att/AI_Engagement_within_Sony_Group.pdf [Accessed: 04.01.2020])

[3] Exemplified by the *Ethics Guidelines for Trustworthy AI* prepared by the European High-Level Expert Group on AI, https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines[accessed: 29.07.2020]

[4] http://harmonious-ai.org/[Accessed: 29.07.2020]

problem with privacy is not primarily the possibility of causing harm but of others showing restraint and not inappropriately infringing on our personal lives. If we walk down the street and, through a window, see a person changing clothes and currently being naked, we tactfully turn away to protect their privacy and to avoid offense or embarrassment. This behavior has been coined "tactful inattention" by the sociologist Erving Goffman. Similarly, a tactful person would not ask a stranger for personal information. Having Takt means to be able to show restraint and not doing something which one theoretically could do out of consideration of another person's feelings. Applied to data-hungry AI systems this would equate to not greedily collecting every bit of data which they are technically able to collect. To give a current example, a computer vision system for detecting if pandemic-related social distancing is upheld in public spaces should either not be built at all or show tactful inattention to the people's identities, genders, cultural background, and other personal information, which is technically quite easy by only using face detection, but not face recognition or any other type of recognition.

- Harmony: "For which tasks should the system not be used? When should it remain silent?" (tactful restraint) "How should the system interact with humans to achieve smooth interactions, avoid causing offense and positively mediate human-human interactions?" (tactful interaction and mediation) "Which information should the system not ask for, record, extract or share?" (tactful privacy)

2) It is not already subsumed by one of the other core principles. The original four principles of bioethics have been introduced in Beauchamp's and Childress' seminal work as "general guidelines for the formulation of more specific rules" and as "four clusters of moral principles" (Beauchamp et al. 2001, p. 13). As such, they constitute a layer of a hierarchy in which each of them subsumes other principles. They serve as denominations for clusters of moral principles which are grouped together based on conceptual vicinity (similar to unsupervised clustering in machine learning based on a distance metric). The principle of justice for example includes the sub-principles of fairness and solidarity. If harmony was just another sub-principle of one of the four core principles, then it should be either derivable from one of them or be conceptually very close.

In the early times of medical ethics, the principles of beneficence and nonmaleficence were regarded as the most important principles as they represent the physician's primary obligations, the desired consequences of his or her actions (curing their patients and not causing harmful side-effects). Justice and especially autonomy rose to prominence only in the last decades when the means of our actions became more important, not just their ends. This focus on the means is strongly reflected in the additional principle of explicability, as having an explainable (and understandable) AI (XAI) or a method for assigning responsibility does not directly influence *what* outcome an intelligent system's actions have, but *how* this result is generated and can be understood. Tactful harmony as we introduced it in this paper is similarly focused on means. It deals with balanced and considerate interactions between agents and less with results. Thus, harmony cannot be subsumed under either beneficence nor non-maleficence. Neither can it be subsumed under justice, as it does not deal with norms of distributing benefits, risks, and costs. The skill of tactful restraint in interactions can give others the room to make their own autonomous decisions, but this is rather a positive side effect next to the primary goal of smoothing the interaction. Thus, tactful harmony is different from the principle of autonomy. As discussed in Section 2, tactful behavior is an embodied, tacit skill which can be communicated only incompletely. As a result, harmony is different from explicability.

3) It does not itself subsume the other principles. Harmony and its herald Takt do not form an ethical super-principle which might subsume and thus replace the other principles of AI ethics. This has already been shown in the last section by discussing the differences between tactful harmony and the other five core principles. Furthermore, harmony in interactions as introduced in this paper has its limitations and cannot serve by itself as a guide towards all which is ethically desirable. Its biggest limitation might be the tendency to avoid conflict. At times, open conflict can be beneficial and a first step towards the resolution of a broader problem. Furthermore, not every offense should be avoided through Takt if it comes at the expense of human rights. As Heyd puts it “Morality protects human interest and rights; tact protects ‘only’ human feelings; and the concern of politeness is merely aesthetic, the style and harmony of everyday life.” (Heyd 1995).

In summary, harmony and tact introduce the valuable perspectives of interaction between AI systems and humans, as well as emotional consideration and awareness of situational particularity. Additionally, as an ethical core principle, it represents issues of privacy and culture-specificity better than any of the existing core principles. Since we have shown that it is neither a sub-category of the other core principles of AI ethics, nor a super-category, we conclude that it would be valuable to add tactful harmony as a sixth core principle.

6 Conclusion

The goal of this paper was to invite a discussion about the new perspective that the concept of harmony could bring the discourse about ethical AI. Harmony is especially valued in East Asian countries and cultures such as China and Japan, but we have shown that the idea was also important to the ancient Greeks and to other influential Western thinkers such as Da Vinci and Kepler. Our synopsis of the different use contexts of harmony revealed that the concept is, at its core, about well-balanced relationships and interactions which give rise to order, balance, and esthetically pleasing phenomena. The relationships of interest are both between the particular and the general. The mediator between these opposites is Takt which is an ability to act thoughtfully and sensible according to the specific situation and to put things into proportion and order. In the sphere of society and human interactions such as education, Takt enables us to perceive each other’s mental and emotional states and act with consideration to avoid conflict and offense.

We have argued that the central challenge of building harmonizing AI is not only to make intelligent systems tactful, but also to design and use them tactfully. For an AI to become tactful, it needs to be able to have advanced sensitivity to the specific contexts which it is in and their social and ethical implications. Furthermore, tactful AI must have some capability of approximately inferring the emotional and cognitive states of people with whom they are interacting and a model on the effects that different possible actions might have on humans. On this dimension of building concordance, AI systems applied in mediating human interaction such as in social networks need to find a balance between connecting people of the same opinion and worldview (unity) with those that hold different opinions and worldviews (diversity) through the choice of which posts and news are shown to them. Intelligent systems applied to personalized news, and more general, information feeds, are in a similar position of needing to find an appropriate balance between showing results that are directly intended and those that broaden the horizon.

For tactful natural interactions with AI, we believe that the research field of conversational informatics which applies techniques from informatics and especially from machine learning to quantitatively study human behavior in conversations will be of high use. In contrast to chat-bots and current digital assistants, it lays special attention on the tacit dimensions of human activity expressed by non-verbal communication such as subtle gestures. Recent progress in AI has brought forth a whole set of powerful tools based on deep learning for pose recognition, speech recognition, activity recognition, inferring emotional correlates in speech prosody and facial expressions, and other methods of human sensing. While these capabilities open the door for intelligent systems to adapt to our predicted emotional states and to act with consideration, they might also be overused, making people feel tricked by machines or virtual avatars imitating emotions. The border between what is desirable and undesirable is sometimes narrow and depends on the specific application and the people affected. AI researchers will need the ability of *Takt* to choose the appropriate and proportional actions. Tactful use of AI is about finding the proportional balance between underuse and over- or misuse as Floridi et al. have argued in Floridi et al. (2018).

On the harmony dimension of rhythm and pace, we argue for the maxim that intelligent systems should always adapt to the human pace instead of the other way around (human-centered). This applies both to the micro-level of human-technology interaction, e.g., in conversation and collaboration as well as to the macro-level of how humanity perceives itself with respect to artificial intelligence. Narratives that call for the necessity of a race against the machines by means of technological enhancement should be treated cautiously. Not because this scientific direction would be wrong in principle, but because the framing of humans needing to keep pace with AI and thus be pushed by it instead of choosing autonomously how to self-actualize and flourish is misguided.

In conclusion, we propose the addition of the principle of harmony to the set of core ethical principles for AI systems. Since the concept of harmony is especially prevalent in Asian cultures such as China and Japan, including it into the canon of central goals and values for ethical AI systems could lead the way towards a stronger inclusion of Eastern perspectives into the currently Western-centric field of AI ethics and yield a more generalizable understanding of how we can apply AI for social good in our globalized society.

Acknowledgment This work was supported by a joint scholarship from RIKEN, Japan, and the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes). We would like to thank both institutions for their support. Furthermore, we would like to thank our reviewers for their excellent comments and highly appreciated feedback.

Funding Open Access funding provided by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommonshorg/licenses/by/4.0/.

References

Aston, W.G. (1896). –Nihongi: Chronicles of Japan from the earliest times to AD 697: translated from the original Chinese and Japanese. The Japan Society, London.

Beauchamp, T.L., Childress, J.F., et al. (2001). *Principles of biomedical ethics*. London: Oxford University Press.

Cooley, M. (1987). *Bee or architect: The human price of technology (revised edition, first edition published 1980)*. London: Hogarth.

Di Dio, C., Macaluso, E., Rizzolatti, G. (2007). The golden beauty: Brain response to classical and renaissance sculptures. *PloS one*, *2*(11), e1201.

Floridi, L. (2013). *The ethics of information*. London: Oxford University Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society, Harvard Data Science Review.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al. (2018). AI4People—an ethical framework for a good Ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

Floridi, L., & Sanders, J.W. (2005). Internet ethics: The constructionist values of homo poieticus. In *The impact of the internet on our moral lives* (pp. 195–214).

French, H.W. (2001). Hypothesis: Science gap. cause: Japan's ways. https://www.nytimes.com/2001/08/07/world/hypothesis-science-gap-cause-japan-s-ways.html.

Gal, D. (2019). Perspectives and approaches in Ai ethics: East Asia. In *Oxford handbook of ethics of artificial intelligence*. Oxford University Press. Forthcoming.

Gladden, M.E. (2019). Who will be the members of society 5.0? Towards an anthropology of technologically posthumanized future societies. *Social Sciences*, *8*(5), 148.

Heyd, D. (1995). Tact: Sense, sensitivity, and virtue. *Inquiry*, *38*(3), 217–231.

Huang, H.-H., & Nishida, T. (2013). Evaluating a virtual agent who responses attentively to multiple players in a quiz game. *Information and Media Technologies*, *8*(1), 81–96.

Ihde, D. (1990). Technology and the lifeworld: From garden to earth.

Iqbal, T., & Riek, L.D. (2017). Coordination dynamics in multihuman multirobot teams. *IEEE Robotics and Automation Letters*, *2*(3), 1712–1717.

Ishiguro, H. (2016). Transmitting human presence through portable teleoperated androids: A minimal design approach.

Jung, M.F., Martelaro, N., Hinds, P.J. (2015). Using robots to moderate team conflict: The case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE international conference on human-robot interaction* (pp. 229–236).

Konishi, E., Yahiro, M., Nakajima, N., Ono, M. (2009). The japanese value of harmony and nursing ethics. *Nursing Ethics*, *16*(5), 625–636.

Kramer, E.M., & Ikeda, R. (1997). What is a "Japanese"?: Culture, diversity, and social harmony in japan. Postmodernism and Race 79–102.

Maris, M., & Boeckhorst, R.ené. (1996). Exploiting physical constraints: Heap formation through behavioral error in a group of robots. In *Proceedings of IEEE/RSJ international conference on intelligent robots and systems. IROS'96*, (Vol. 3 pp. 1655–1660): IEEE.

Naukkarinen, O. (2014). Everyday aesthetic practices, ethics and tact. *Aisthesis. Pratiche, linguaggi e saperi dell'estetico*, *7*(1), 23–44.

Nishida, T. (2008). *Conversational informatics: an engineering approach* Vol. 9. New York: Wiley.

Nishida, T. (2016). *Human-harmonized information technology, volume 1: vertical impact*. Berlin: Springer.

Nishida, T. (2017). *Human-harmonized information technology, volume 2: horizontal expansion*. Berlin: Springer.

Nishida, T., Nakazawa, A., Ohmoto, Y., Mohammad, Y. (2014). *Conversational informatics: a data-intensive approach with emphasis on nonverbal communication*. Berlin: Springer.

Government of Japan (2016). The fifth science and technology basic plan. Provisional translation. https://www8.cao.go.jp/cstp/english/society5_0/index.html. PDF available online: https://www8.cao.go.jp/cstp/english/basic/5thbasicplan.pdf (accessed on 24 September 2019).

Rahwan, I., Crandall, J.W., Bonnefon, J.-F. (2020). Intelligent machines as social catalysts. In *Proceedings of the National Academy of Sciences*.

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al. (2019). Tackling climate change with machine learning. arXiv:1906.05433.

Stakhov, A. (2009). The mathematics of harmony: From Euclid to contemporary mathematics and computer science, Vol. 22, World Scientific, Singapore.

Strubell, E., Ganesh, A., McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv:1906.02243.

Suzuki, S. (2008). Takt als Medium. Überlegungen zum Takt Begriff von J. F. Herbart, Internationale Zeitschrift für Historische Anthropologie,17 (1).

Suzuki, S. (2010). Takt in modern education waxmann.

Suzuki, S. (2012). The Kyoto School and J.F. Herbart. In *Education and the Kyoto School of Philosophy* (pp. 41–53): Springer.

Suzuki, S. (2014). Takt. In *Handbuch Pädagogische Anthropologie* (pp. 295–301): Springer.

Suzuki, S. (2019). Etoku (会得) and rhythms of nature. In Resina, J.R., & Wulf, C. (Eds.) *Repetition, recurrence, returns* (pp. 131–146): Lexington Books.

Traeger, M.L., Sebo, S.S., Jung, M., Scassellati, B., Christakis, N.A. (2020). Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences*, *117*(12), 6370–6375.

Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. University Park: Penn State Press.

Verbeek, P.-P. (2009). The moral relevance of technological artifacts. In *Evaluating new technologies* (pp. 63–77): Springer.

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. Chiacago: University of Chicago Press.

Verbeek, P.-P. (2015). Toward a theory of technological mediation. In *Technoscience and postphenomenology:, the Manhattan* (pp. 189–204).

Wada, K., & Shibata, T. (2007). Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, *23*(5), 972–980.

Wierzbicka, A. (1991). Japanese key words and core cultural values. *Language in Society*, *20*(3), 333–385.

Wong, P.-H. (2012). Dao, harmony and personhood: Towards a confucian ethics of technology. *Philosophy and Technology*, *25*(1), 67–86.

Yu, K.P. (2010). The Confucian conception of harmony. *Governance for Harmony in Asia and Beyond*, *7*, 15.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.