2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR) | 978-1-7281-7463-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/AIVR50618.2020.00032

Introducing the concept of ikigai to the ethics of AI and of human enhancements

Soenke Ziesche
Independent researcher
Delhi
India
soenke.ziesche@gmail.com

Roman Yampolskiy
Speed School of Engineering
University of Louisville
USA
roman.yampolskiy@louisville.edu

***Abstract*— It has been shown that an important criterion for human happiness and longevity is what is expressed by the Japanese concept of ikigai, which means "reason or purpose to live". In the course of their lives humans usually search for their individual ikigai, ideally find it and hence devote time to it. As it is widely expected that both AI and XR will be increasingly disruptive of our known daily time use schedule, this will likely also have an impact on the space of potential ikigai. Since ikigai constitutes a vital component of the lives of humans, these consequences for ikigai have to be examined towards both ethical human enhancement as well as ikigai-friendly AI. In this paper the term "i-risk" is introduced for undesirable scenarios in which humans and potentially also other minds are deprived of the pursuit of their individual ikigai. This paper outlines ikigai-related challenges as well as desiderata for the three categories XR/human enhancement, AI safety and AI welfare.**

***Index Terms*— ikigai, time use, i-risks, ethics, XR, human enhancement, AI safety, AI affordances, AI effect, AI welfare**

I. INTRODUCTION TO IKIGAI AND TIME USE RESEARCH

Ikigai refers to a Japanese concept, which can be translated as "reason or purpose to live". It comprises those activities of life, which provide satisfaction and meaning.

Obviously, other cultures and philosophers have contemplated similar questions and concepts, such as the French "raison d'être" or the Greek "eudaimonia", yet for consistency we use the term ikigai here, which captures what we intend to convey.

At the same time, in recent years Western authors have "discovered" the topic of ikigai, which resulted in various popular scientific publications, e.g. [1], and which is not further elaborated here since these approaches are often based on misconceptions and limited, e.g. by linking ikigai to somebody's professional career only. Therefore, the original works about ikigai are more appropriate as well as sufficient for us.

As Kamiya, one of the earliest ikigai theorists, pointed out, ikigai has various connotations [2]. An important distinction is described by the two aspects ikigai kan and ikigai taishō, e.g. [3]. Ikigai kan expresses feelings of satisfaction, well-being and a life worth living, thus it is a state of mind, while ikigai taishō comprises activities, experiences and circumstances, which lead to such feelings, thus it is rather a process. An important feature in Japanese culture is for individuals to reflect extensively towards a suitable ikigai taishō. Once a person has identified a personal ikigai taishō, she or he embraces the associated activities towards his or her subjective ikigai kan.

Examples for ikigai taishō include family life, hobbies, professional occupations and other social activities, especially, but not only such activities, which help others. Ikigai taishō can comprise a variety of activities from small everyday rituals to the pursuit of complex goals.

As we will specify below, one of the reasons for the interest in the concept of ikigai in recent years is that studies have revealed a correlation between pursuing an individual ikigai and health, hence longevity. In other words, if humans have nothing what makes their life worth living then this leads not only to unhappiness, but impacts also their health, thus their life expectancy. Therefore, it is desirable to maximize the number of humans (or sentient digital minds in general, see below) who have found an ikigai and allow them to pursue it.

In this regard we also have to explore likely disruptions by upcoming technologies such as AI and human enhancements, e.g. through extended reality (XR). XR refers to any combinations of real and virtual environments, which are increasingly refined [4]. The topic of ikigai has so far not been linked to these technologies. Below we introduce visionary contemplations of XR, which look at the well-being and happiness of humans. Yet, this is the static ikigai kan part, while the necessary ikigai taishō part (towards ikigai kan) appears to have been neglected, which would be the related day-to-day activities of humans in times of AI and human enhancements.

Due to both the significance of finding and pursuing an ikigai for humans as well as the expected disruption of our daily lives by AI and XR, it is argued here that the facilitation of ikigai should be an indispensable part of ethics for AI and XR. Two dimensions are foreseeable:

- Currently common ikigai activities may disappear, e.g. certain professional occupations, which are by many treated as their ikigai.
- New, currently unknown ikigai activities may appear owing to new possibilities through AI and XR.

If this challenge is not prudently and timely addressed we foresee what we call "risk of ikigai loss" or "i-risk". This can be seen as another level of the terms x-risks and

978-1-7281-7463-1/20/$31.00 ©2020 IEEE
DOI 10.1109/AIVR50618.2020.00032

s-risks introduced by others before. Bostrom coined and described x-risks or existential risks, which may lead to annihilation of life on earth [5]. Althaus and Gloor added the category of s-risks or suffering risks, which may increase suffering drastically [6]. While x-risks and s-risks are critical criteria to define undesirable scenarios, we argue that in addition scenarios are to be avoided, in which humans are alive and not suffering per se, but are devoid of any ikigai. This could be a world where ikigai activities have vanished or humans are prevented to find and pursue their ikigai. Risks, which may lead to such scenarios, we call i-risks.

In order to address i-risks this article aims to introduce relevant synergies between the traditional concept of ikigai and the fields AI safety, human enhancement as well as AI welfare. For a scientific foundation we introduce first the two concepts of metrics for ikigai and time use research.

A. *Metrics for ikigai*

It can be objected that ikigai is an intangible concept; hence in order to scientifically discuss ikigai, metrics are required. We look at two dimensions of measurements:

- To verify that having found and pursuing an ikigai is correlated with health and longevity.
- To examine whether a suitable ikigai taishō for individual people can be calculated.

For the first topic research has been conducted not only regarding ikigai in Japan, but also in general regarding the purpose of life, e.g. [7]. More specifically, for example, Sone et al., Tanno et al. and Tomioka et al. concluded that for Japanese people having an ikigai is associated with health and longevity [8]-[10]. Since ikigai can have diverse manifestations, lists have been developed to assess multidimensional aspects of ikigai instead of merely asking subjects whether they found an ikigai. For example, Imai proposed "ikigai-9", which comprises nine sub-themes of ikigai [11].

For the second topic no existing specific research has been found. However, for example, Helliwell et al. analyze happiness and wellbeing in the annual "World Happiness Report". They also look at big data approaches and came to the conclusion "that Big Data is increasing the ability of researchers, governments, companies, and other entities to measure and predict the well-being and the inner life of individuals." (p. 119)[12].

While this is more associated with the static ikigai kan, we propose that big data could be also used to calculate related ikigai activities, i.e. ikigai taishō[1]. There are different categories of big data. For happiness measurement especially user-generated digital content, e.g. on social media, is relevant, from which wellbeing can often be directly derived. Other big data may require more, usually AI supported, analysis to uncover information about individuals' tastes and choices. Overall, this may also lead to insights about the individuals' pursuit of ikigai activities. Also a comparison with dating apps and the associated algorithms may be suitable: Algorithms could be developed and through machine learning improved, which suggest potentially matching ikigai activities to individuals instead of potentially matching partners.

We also note that big data carry significant privacy concerns in general and for ikigai research in particular since the quest for ikigai is traditionally considered a private reflection. However, the benefits and drawbacks have to be weighed: If metrics and methods can be developed to find ikigai for people, thus to increase their happiness and health, this may outweigh the privacy issues.

B. *Time use research*

Another relevant field of study for this article is time use research, which examines how humans on average allocate their time to certain activities. The main methodology in this regard are statistical time use surveys, which are being conducted periodically in different countries. For example Charmes provides a comparative analysis of 102 time use surveys carried out in 65 countries, which reveals significant differences of time use between different countries as well as between different sexes [14].

In order to allow comparison between time use surveys standardized categories of activities are required. One leading example is the International Classification of Activities for Time-Use Statistics (ICATUS) from 2016, which has the following nine categories on the top level and various further sub-categories:

- Employment and related activities
- Production of goods for own final use
- Unpaid domestic services for household and family members
- Unpaid caregiving services for household and family members
- Unpaid volunteer, trainee and other unpaid work
- Learning
- Socializing and communication, community participation and religious practice
- Culture, leisure, mass media and sports practices
- Self-care and maintenance

The categories are mostly self-explanatory, hence we just mention that sleeping is part of self-care and maintenance. An older, but also relevant classification is the one by Ås, which distinguishes the following four categories [15]:

- Necessary time: time spent on activities for physiological needs
- Contracted time: time that human beings spend to fulfill the contracts that they have made
- Committed time: time committed to fulfill social responsibilities
- Free time: the residual time left after performing contracted, committed and necessary time

The last relevant classification introduced here is between primary and secondary time, which allows examining multitasking. In such situations primary time

[1] See also the recently proposed scheme of Augmented Utilitarianism in XR contexts [13].

refers to the main activity and secondary time to the side activity.

For our purposes it is relevant to investigate 1) under which time use categories ikigai activities fall, 2) how much time people can devote to their individually ikigai activities, and 3) how this may change in the future due to developments in AI and XR? As noted before in particular the first topic depends on individual preferences as, for example, for some people their employment is their ikigai, while others pursue their ikigai during social or leisure activities. And for the second topic again the differences between countries, in particular between developing and developed countries, have to be highlighted. People may have an ikigai, but may have hardly any time to pursue it, one out of many examples being women and girls in developing countries, who spend significant amounts of time on collecting water every day.

Time use research appears to have so far not been linked to ikigai. Gershuny examines relations between time use and wellbeing without referring to the distinction between ikigai kan and ikigai taishō [16]. For example Gershuny compares between countries as well as sexes enjoyment ratings on a 0-10 scale for certain activities [16].

The introduced metrics for ikigai as well as time use research are revisited below when we examine how it could be avoided that human enhancement or AI lead to an increase of i-risks.

II. IKIGAI AND XR/HUMAN ENHANCEMENT

A. Introduction

We define enhanced humans here as humans who still have biological parts, but whose capacities regarding health span, emotion and cognition (vastly) exceed current humans, for example by means of XR technologies or brain–computer interfaces. The field is progressing and applications have, for example, focused on entertainment [17], education [18], healthcare [19] manufacturing [20] as well as ethical self-assessment [21].

Usually enhancements are described as beneficial for humans in various aspects, thus desirable, e.g. [22]. Especially the potential achievement of enduring bliss is often mentioned and motivates research in the field, e.g. [23]. Yampolskiy describes a particular scenario, which also addresses the AI value alignment problem (see further below), whereby personalized XR simulations for humans are created as Individual Simulated Universes tailored for individual values and happiness [24].

B. Challenge

As just introduced the cited aim of XR or further enhancements is often for the concerned humans to obtain bliss as opposed to suffering, e.g. [23]. Yet, here it is critical to reiterate the distinction between s-risks and i-risks: We argue that the absence of suffering is not a sufficient criterion for bliss, but bliss requires ikigai too, i.e. reason or purpose to live. In other words, scenarios with no or minimal suffering have no or low s-risks, but there may still be significant i-risks, which would concern non-suffering humans devoid of an ikigai. Non-suffering humans without ikigai exist also in our current world, yet we highlight that this situation may exacerbate as enhancement-related disruption may lead to the disappearance of traditional ikigai activities. Examples include the abolishment of professional occupations, which were considered as ikigai by some, such as teachers who could be replaced by XR-enabled education.

Regarding day-to-day activities in XR and enhancement scenarios discussions have so far mostly focused on avoidance of boredom and availability of sufficient novelties for the concerned humans to experience [25]-[27]. Such deliberations about non-boring activities are relevant, especially for anticipated very long lifetimes owing to enhancements, yet do not address exactly the issue of ikigai. In other words, an environment with manifold opportunities is helpful to find a purpose of life, but another step towards ikigai is required.

This challenge can be more specifically examined through time use statistics: Since the space of possible activities for enhanced humans will be by definition very different from current day-to-day activities their time use schedule will be modified. Table **1** illustrates how the relative time-share may change for the four categories necessary time, contracted time, committed time and free time introduced by Ås [15]:

	Necessary time	**Contracted time**	**Committed time**	**Free time**
Examples while in enhanced/ XR modus	None, because by definition necessary time is spent on physiologi-cal, i.e.non-enhanced /XR needs.	New professions enabled through enhancement /XR.	Social interactions in XR/enhanced environment.	Entertainment and education in XR/enhanced environment.
Time-share enhanced/ XR modus compared with non-enhanced/ XR modus	Same.	Smaller, especially if professional occupations become obsolete.	Potentially larger. May take over share of previously contracted time-share.	Likely larger. Takes over share of previously contracted time-share.

Table 1. Anticipated time-share changes due to XR and enhancements

In addition to the relative shifts of time use it is also likely that humans will overall have more time at hand for the following two reasons:

1. Enhanced cognition could enable more efficient multitasking, i.e. formerly consecutive activities may be conducted in parallel during primary and secondary time.
2. Enhancements are likely to increase the life span as well as health span.

While it will be up to future time use research to measure how exactly enhanced humans spend their time, for us the focus is on the subset of ikigai taishō activities. Scenarios are imaginable in both directions that in a future world of XR and enhancements the space of ikigai taishō activities will be restricted, e.g. through abolishment of

professional occupations by enhancements, which were nevertheless considered as ikigai by some, as mentioned above, or, perhaps more likely, that the space of ikigai taishō activities will be enlarged, i.e. with the availability of more time as well as more options, especially in the categories committed and free time.

C. Ikigai-related desiderata

Ikigai research should be a critical component for framing ethical guidelines for content creators, developers, distributors as well as users of XR. The minimal goal should be that neither the available options nor the daily time for ikigai taishō activities would be shortened due to XR and further enhancements, while ideally XR and further enhancements provide for more options as well as even longer daily periods of ikigai taishō activities. However, the specifics of these ikigai taishō activities may be in current pre-enhancement times still unknown, yet as meaningful as well as satisfying.

Therefore, ethics for XR should carefully consider future time use from an ikigai-optimization point of view, possibly integrated in context-sensitive frameworks such as Augmented Utilitarianism [13]. This should be in cooperation with time use researchers based on their established classifications, such as ICATUS, which may have to be extended. Ethics should especially consider the scenario that (large) parts of the known time use schedule might disappear; hence the emerging vacuum has to be carefully filled. Recalling the current gap between inhabitants of developing and of developed countries concerning available time for pursuing ikigai activities, enhancements and the related ethics should in particular establish more equality in this regard. In fact, XR and enhancements may allow some humans for the first time to pursue ikigai activities as they may have been before entirely occupied by necessary and contracted time [15]. While activities during contracted time are actually considered by some as ikigai, the larger time-share of committed and especially of free time means also more freedom in finding a personal ikigai.

In order to design and also to structure the unprecedented, largely artificial range of possibilities in XR and enhanced environments or even individual universes new occupations will be required. For example, Yampolskiy suggested in this regard the profession of a Universe Designer [28]. While preferably some or many traditional ikigai activities should continue to remain available in XR and enhancement environments and potentially for longer time periods per day, recalling the risk of boredom and depression over long lifetimes, further novel options should be explored too. Since the potential of ikigai activities for enhanced minds is both likely to be very different from what we know and impossible to forecast at this stage, an "Ikigai Designer" may be needed, similarly to Yampolskiy's Universe Designer.

An advantage of such digital environments is that (big) data about activities, time use as well as states of minds from enhanced humans can be much easier retrieved, which may facilitate ikigai calculation, i.e. proven and tested ikigais from enhanced humans could be recommended to others.

In summary, the potential opportunities, thus desiderata, provided by XR and supported by dedicated ethics are:

1. Humans will likely find new formerly unknown ikigai taishō.
2. Humans will likely have more freedom for ikigai activities because they may have less contracted time and more committed as well as more free time.
3. Humans will likely have more time and for ikigai activities because they may conduct non-ikigai activities in a more time-efficient manner and they may live longer.
4. In addition to more time and health, XR will likely furnish other resources, which are required for many traditional ikigai activities, but tend to decline for ageing people [29], such as social networks.

III. IKIGAI AND AI SAFETY

A. Introduction

While a variety of problems towards AI safety has been identified, e.g. [30], the focus here in connection with ethics for XR and further enhancements is on one of the most prominent ones, which is the value alignment problem. Yudkowsky and Bostrom described this problem and also pioneered this field [31], [32]. The basic question is how to cause an AI to pursue goals and values, which are aligned with human goals and values. A failure in solving this problem may constitute an existential threat to humanity since there is no reason to assume that an AI will turn out to be value-aligned with humans without prior arrangements. The value alignment problem can be divided in sub-problems such as to agree on common human values, to precisely specify these values in a machine-understandable way and others. These sub-problems, thus the value alignment problem as a whole, have been proven to be very hard despite significant work in recent years, e.g. [33], [34].

B. Challenge

Acknowledging that AI value alignment is indispensable as well as very hard to achieve, we argue that it is equally important that any AI must not impede the ikigai of individual humans because ikigai is vital for humans as shown above. To be more precise, we claim not only that ikigai is neglected by current AI value alignment research, but also that achievement of AI value alignment would not necessarily entail that humans have and can pursue individual ikigai. Therefore, not taking ikigai into consideration for AI value alignment research would constitute an i-risk, which we seek to avoid.

We illustrate the differences between values and ikigai in Table 2 as well as by an example scenario:

Values	Condition	Discrete	Common, shared and mutual	Should be aligned.
Ikigai (taishō)	Activity	Holistic, big picture	Individual and diverse	Should be aligned (in the sense that they do not violate other people's values or impede other people's ikigai).

Table 2. Comparison between values and ikigai

The following scenario provides an example when value alignment does not imply freedom of ikigai: It is assumed that there is an AI, which is aligned with human values, but massively restricts the space of potential ikigai activities to e.g. for everyone to play musical instruments only (perhaps because the AI requires other matter as resources for its goals or the qualia of music creates pleasure for the AI). Playing musical instruments is indeed an ikigai for some people, but not for all, thus in this scenario value alignment is achieved (as per assumption), while the pursuit of individual ikigai under the regime of such an AI is not accomplished. Therefore, lack of freedom of ikigai could be also taken as indicator for perverse instantiation issues [32].

Again we stress that our focus is on day-to-day activities of humans in future times largely affected by powerful AIs. Since it has been shown that the pursuit of individual ikigai activities are linked to happiness and health and since the penetration of daily lives by AIs will have an impact of the space of possible activities, many humans would likely have to adjust or change their usual ikigai, hence research on future ikigai activities is critical. Similar to XR and further enhancements, AIs are likely to influence the space of possible activities in both directions: Ideally due to support by AIs there may be less currently indispensable activities for humans to address their basic needs, including employment, which often constituted ikigai as well, but there could be also new so far unknown fulfilling activities.

C. Ikigai-related desiderata

Similar to the term "friendly AI" [31] an "ikigai-friendly AI" is desirable, of which three types are conceivable:

1. The AI does not prevent humans from searching, finding and pursuing their individual ikigai.

   Exceptional cases would be if a certain ikigai were in conflict with Omohundro's drives [35]. Examples are if the ikigai of an individual is to destroy AIs or if the ikigai of an individual and the AI both require the same resources.

2. In addition to the features of type 1, the AI has the knowledge how to help humans to find their individual ikigai and teaches it to individuals upon request.

   A scenario would be that the AI gains the knowledge what would be fulfilling ikigai activities for individuals through big data analysis as introduced above. While, as mentioned, the quest for ikigai is traditionally considered a private reflection, support through an AI may be more sought-after, since the advent of powerful AIs may have disrupted and overwhelmed people's lives due to AI-related unemployment or other consequences.

   This may include currently unknown ikigai activities; either because humans have never conceived them or these activities were only enabled by the AI, such as new kinds of hobbies or social interaction with artificial minds. These new ikigai activities could be more alien than innovative ikigais in XR environments due to the superiority of the AI and could be referred to as AI affordances [36], i.e. actions only made possible by an AI environment. This could be called AI effect, similar to the observer effect in physics. In other words, because of the AI we may revise our whole concept of ikigai.

3. In addition to the features of types 1 and 2, the AI manipulates humans towards a specific ikigai. In that way an AI environment may change our values and goals in an unforeseeable, but positive manner. For example, an ikigai-friendly AI is aware what is best for humanity, i.e. environmentalism, healthy lifestyle and nutrition, inclusion, overall as well as gender equality, anti-racism and anti-speciesism, may then come up with supporting ikigai activities in this regard and indoctrinates humans to consider a convergent ikigai, which they would not have done otherwise.

   While this is a positive scenario of a manipulative AI, also negative, thus not ikigai-friendly, options are conceivable, which is that the AI manipulates or applies wireheading to humans, e.g. [37], so that the affected humans consider activities, which support the AI goals, but which may include seemingly dull undertakings, as ikigai. An example would be that the AI seduces humans to produce paperclips in case the AI is a paperclip maximizer [38]. The difference to similar enslavement scenarios would be that the affected humans would not suffer, but be under the impression to pursue their reason for being[2].

An additional feature of all three types of "ikigai-friendly AI" could be that it recognizes non-ikigai activities within individual time use schedules of humans and looks at ways how to take them over in order for the human to free time for her or his ikigai activities.

In summary, the potential opportunities provided by an ikigai-friendly AI are the same as mentioned above for XR, which includes potentially even more novel ikigai activities owing to AI affordances as well as more time for ikigai activities. However, there is also the risk of an "ikigai-unfriendly AI", which may not allow humans to pursue ikigai or which may wirehead humans in order treat activities as ikigai that benefit the AI's goals, but would usually not be entertained by humans.

[2] See also for similar thoughts regarding AI-based religions [39].

IV. IKIGAI AND AI WELFARE

A. Introduction

It has been claimed that there is a non-zero probability that sentient digital beings not only exist, but also that they may, at least temporarily, suffer, which may include AIs, e.g. [40]-[42]. Therefore, Bostrom et al. developed a policy desideratum concerning the interests of sentient digital minds, which states "AI is governed in such a way that maltreatment of sentient digital minds is avoided or minimized" (p. 18) [41]. In follow up to this desideratum Ziesche and Yampolskiy proposed a new field "AI welfare science" and offered recommendations for necessary activities towards the wellbeing of sentient digital minds as well as comprehensive antispeciesism [42]. Certain not yet existing but in the future theoretically possible AIs have been categorized by Aliman et al. as Type II systems, "systems with a scientifically plausible ability to act independently, intentionally, deliberately and consciously and to craft explanations." (p. 2) (while Type I systems are defined as the complement of Type II systems with all present-day AIs being of Type I) [43].

B. Challenge

We distinguish two potential ikigai-related challenges for sentient digital beings:

1. If there is a non-zero probability that sentient digital beings exist, it could be argued that there is also a non-zero probability that sentient digital beings not only have the concept of ikigai, but also the longing to pursue an ikigai. One of the defining features of sentient digital beings is the assumed capacity to perceive qualia. Ziesche and Yampolskiy look in their introductory paper mostly at suffering through perception of qualia of pain [42]. However, also other concerns could impact the wellbeing of humans as well as potentially the wellbeing of sentient digital minds. Bostrom et al. note as examples for other types of conceivable mistreatment of sentient digital minds restrictions to their autonomy, creativity and self-expression [41].

These restrictions can also be interpreted as lack of freedom or lack of opportunities for sentient digital minds to pursue an individual ikigai. In other words, concerned sentient digital minds without ikigai may suffer even if they do not perceive qualia of pain. This leads to a second dimension of the i-risks introduced above. On the one hand, if not carefully handled, many humans may lose their ikigai in times of AI and XR, as described above. On the other hand, AI and XR may produce (a high number of) sentient digital minds, which have the concept as well as the longing for ikigai, but may be prevented from finding and pursuing it. In other words, AI and XR may constitute i-risks not only because they may cause the number of humans without ikigai to increase, but also because they may cause the number of other ikigai-requiring digital minds to increase.

2. As introduced, some ikigais may involve social activities, while other do not. We look here at the first category, the significance of which for example Fukuzawa et al. highlighted by showing the relevance of social networks for ikigai [29].

XR technologies provide opportunities in this regard, but also largely neglected risks. Not only many games, but also other XR environments are based on interacting with (avatars of) other people. However, an integral part of social activities in XR are, apart from avatars of other humans, non-player characters (NPCs). And if indeed XR environments prove to provide innovative ikigais based on social activities, a large number of NPCs may be involved. Since NPCs could be sentient too [44] and hence could have a moral status, Ziesche and Yampolskiy argue, that policies are required for NPCs towards ethical human enhancement and antispeciesism [45]. Therefore, even if the above claim regarding i-risks for sentient digital minds is not true, XR environments, which support ikigais based on social activities, may increase suffering, i.e. pose an s-risk [6].

C. Ikigai-related desiderata

Ziesche and Yampolskiy developed recommendations for AI welfare policies regarding suffering of sentient digital minds, yet, as mentioned, initially limited to suffering through qualia of pain [42]. These should be also applicable for XR and AI environments, which have the purpose for humans to pursue their ikigai. Moreover, we propose to complement these recommendations as follows to also address suffering of sentient digital minds because of lack of ikigai:

- Initiate research on AI welfare science to develop methods to measure if sentient digital minds long for an ikigai.
- Potentially initiate research on AI welfare science to develop methods how individual ikigais could be found for sentient digital minds and how the pursuit of these individual ikigais by the sentient digital minds could be ensured.

Ziesche and Yampolskiy outline preliminary considerations and challenges regarding methodologies such as self-reporting as well as observation through functional and behavioral parameters [42].

Moreover, Ziesche and Yampolskiy proposed that humans are deterred from harming minds in other substrates by mandatorily equipping the humans with sensory perception of potential pain of the NPCs in order for the humans to understand when to stop activities because they cause the suffering [45]. Similarly, it could be made compulsory that humans in AI and XR environments are notified of frustration of NPCs prevented from finding and pursuing an ikigai, especially if the humans are engaged in their own individual ikigais based on social activities with these NPCs.

Future time use research for sentient digital minds may be useful in order to establish if sentient digital minds have time at all to allocate to ikigai activities. In this regard, one feature to take into account is that sentient digital minds may have a very different subjective rate of time [46].

Two points have to be highlighted:

- This section is largely speculation. Nevertheless, such prolegomena are relevant to reduce potential (massive) s-risks as well as i-risks in AI and XR environments, as initiated by Bostrom et al. [41] and as further specified by Ziesche and Yampolskiy [42].
- Anthropomorphic bias must be taken into account: 1) For other minds ikigai may not be relevant, 2) or they may have ikigais, which appear very alien to us and may be beyond our imagination, 3) or they may have obtained ikigai through wireheading, which may be sufficient and satisfying for them.

V. Summary and Future Work

In this paper we have argued that it is essential to add the topic of ikigai to the fields of ethical AI and human enhancement and to raise awareness for content creators, developers, distributors, and users of XR. In particular, we introduced the following overlooked challenges as well as suitable concepts:

- i-risks concern scenarios, in which human and potentially other minds are not able to pursue their individual ikigai, and constitute a distinct level apart from x-risks and s-risks. We have identified unforeseen challenges for human enhancement, AI safety and AI welfare, which may lead to i-risks, as well as desiderata in this regard and some initial proposals for solutions.
- Aspects that increase i-risks: XR and AI will likely be disruptive in a way that certain traditional ikigais may vanish, especially professional occupations. Also, AIs may restrict the freedom of humans to pursue their ikigai.
- Aspects that reduce i-risks: If handled rightly, XR and (ikigai-friendly) AI may also provide affordances, opportunities and resources for new ikigais. Moreover, XR and AI could ensure more time for ikigai activities, including for those whose time use schedule hardly had slots for it before, i.e. XR and AI could reduce inequality among humans. To further reduce i-risks also potential ikigai needs of sentient digital minds have to be taken into account.
- We also proposed methods how to measure and how to find ikigai for individual people through big data as well as extended time use research towards both XR and AI environments.

In this paper we have provided groundwork, but significant future work can be foreseen:

- For the outlined desiderata detailed specifications and solutions are required.
- Moreover, the topic of ikigai and ethics could be widened to look at overall mental health policies in preparation for expected massive disruptions due to enhancements and AI. Since these disruptions may cause economic pressures as well rapid social change and since the World Health Organization noted[3] that this may harm mental health, a contingency plan and related policies for a potential increase of mental health issues appear to be prudent to set up.

References

[1] García, H. & Miralles, F. Ikigai (2017). *Ikigai: The Japanese Secret to a Long and Happy Life*. Penguin.

[2] Kamiya, M. (1966). Ikigai-ni-tsuite[On 755 ikigai]. Tokyo, Japan: MisuzuShyobou.

[3] Mathews, G. (2008). Finding and keeping purpose in life. *Pursuits of happiness: Well-being in anthropological perspective*, 167.

[4] Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv preprint arXiv:1804.08386*. https://arxiv.org/pdf/1804.08386.pdf (accessed 19 Sep 2020).

[5] Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and technology*, *9*. https://nickbostrom.com/existential/risks.html (accessed 19 Sep 2020).

[6] Althaus, D., & Gloor, L. (2016). Reducing risks of astronomical suffering: a neglected priority. *Foundational Research Institute*. Available online: https://foundational-research.org/reducing-risks-ofastronomical-suffering-a-neglected-priority/ (accessed 19 Sep 2020).

[7] Crumbaugh, J. C., & Maholick, L. T. (1964). An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *Journal of clinical psychology*, *20*(2), 200-207. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.505.6866&rep=rep1&type=pdf (accessed 19 Sep 2020).

[8] Sone, T., Nakaya, N., Ohmori, K., Shimazu, T., Higashiguchi, M., Kakizaki, M., ... & Tsuji, I. (2008). Sense of life worth living (ikigai) and mortality in Japan: Ohsaki Study. *Psychosomatic medicine*, *70*(6), 709-715. https://se-realiser.com/wp-content/uploads/2018/11/Sone-et-al.-Ikigai-and-Mortality-in-Japan-2008.pdf (accessed 19 Sep 2020).

[9] Tanno, K., Sakata, K., Ohsawa, M., Onoda, T., Itai, K., Yaegashi, Y., ... & JACC Study Group. (2009). Associations of ikigai as a positive psychological factor with all-cause mortality and cause-specific mortality among middle-aged and elderly Japanese people: findings from the Japan Collaborative Cohort Study. *Journal of psychosomatic research*, *67*(1), 67-75.

[10] Tomioka, K., Kurumatani, N., & Hosoi, H. (2016). Relationship of having hobbies and a purpose in life with mortality, activities of daily living, and instrumental activities of daily living among community-dwelling elderly adults. *Journal of epidemiology*, *26*(7), 361-370. https://www.jstage.jst.go.jp/article/jea/26/7/26_JE20150153/_pdf (accessed 19 Sep 2020).

[11] Imai, T. (2012). The reliability and validity of a new scale for measuring the concept of Ikigai (Ikigai-9). *[Nihon koshu eisei zasshi] Japanese journal of public health*, *59*(7), 433-439.

[12] Helliwell, J. F., Layard, R., & Sachs, J. D., & (2019). *World happiness report 2019*. https://s3.amazonaws.com/happiness-report/2019/WHR19.pdf (accessed 19 Sep 2020).

[13] Aliman, N.-M., Kester, L. & Werkhoven, P. J. (2019). XR for Augmented Utilitarianism. In *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 283-285.

[14] Charmes, J. (2015). Time use across the world: Findings of a world compilation of time use surveys. *UNDP Human Development Report Office, background Paper, New York*. http://www.hdr.undp.org/sites/default/files/charmes_hdr_2015_final.pdf (accessed 19 Sep 2020).

[3] See: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

[15] Ås, D. (1978). Studies of time-use: problems and prospects. *Acta Sociologica*, *21*(2), 125-141. https://www.timeuse.org/sites/default/files/public/ctur_journal_article/2179/DagfinnAs.pdf (accessed 19 Sep 2020).

[16] Gershuny, J. (2011). Time-use surveys and the measurement of national well-being. *Centre for Time Use Research, University of Oxford, Swansea, UK, Office for National Statistics*. https://www.timeuse.org/sites/ctur/files/public/ctur_report/4486/timeusesurveysandwellbein_tcm77-232153.pdf (accessed 19 Sep 2020).

[17] Von Itzstein, G. S., Billinghurst, M., Smith, R. T., & Thomas, B. H. (2019). *Augmented Reality Entertainment: Taking Gaming Out of the Box*. https://www.researchgate.net/profile/Bruce_Thomas3/publication/318183569_Augmented_Reality_Entertainment_Taking_Gaming_Out_of_the_Box/links/5e92eddf4585150839d65361/Augmented-Reality-Entertainment-Taking-Gaming-Out-of-the-Box.pdf (accessed 24 Oct 2020).

[18] Liu, D., Dede, C., Huang, R., & Richards, J. (Eds.). (2017). *Virtual, augmented, and mixed realities in education*. Singapore: Springer.

[19] Andrews, C., Southworth, M. K., Silva, J. N., & Silva, J. R. (2019). Extended reality in medical practice. *Current treatment options in cardiovascular medicine*, *21*(4), 18. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6919549/ (accessed 24 Oct 2020).

[20] Fast-Berglund, Å., Gong, L., & Li, D. (2018). Testing and validating Extended Reality (xR) technologies in manufacturing. *Procedia Manufacturing*, *25*, 31-38. https://pdfs.semanticscholar.org/316e/03ddb48e56dfd73ce4cbc6a72e059bb2d46e.pdf (accessed 24 Oct 2020).

[21] Aliman, N.-M. & Kester, L. (2019). Extending socio-technological reality for ethics in artificial intelligent systems. In *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 275-282.

[22] Bostrom, N. (2008). Why I want to be a posthuman when I grow up. In *Medical enhancement and posthumanity*. Springer, Dordrecht: 107-136. https://nickbostrom.com/posthuman.pdf (accessed 19 Sep 2020).

[23] Faggella, D. 2019. The Transhuman Transition – What it is and Why it Matters. https://danfaggella.com/transhuman-transition/ (accessed 19 Sep 2020).

[24] Yampolskiy, R. V. (2019). Personal universes: A solution to the multi-agent value alignment problem. arXiv preprint arXiv:1901.01851. https://arxiv.org/pdf/1901.01851.pdf (accessed 19 Sep 2020).

[25] Ziesche, S. & Yampolskiy, R. V. (2016). *Artificial Fun: Mapping Minds to the Space of Fun.* arXiv preprint arXiv:1606.07092. https://arxiv.org/ftp/arxiv/papers/1606/1606.07092.pdf (accessed 19 Sep 2020).

[26] Loosemore, R.P.W. (2014). Qualia Surfing. In R. Blackford & D. Broderick (Eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds*. Chichester: John Wiley & Sons, 231-239.

[27] Yudkowsky, E. (2009). *The Fun Theory Sequence*. http://lesswrong.com/lw/xy/the_fun_theory_sequence/ (accessed 19 Sep 2020).

[28] Yampolskiy, R.V. (2017). Future Jobs - The Universe Designer. *Circus Street*. https://www.circusstreet.com/blog/future-jobs-the-universe-designer (accessed 19 Sep 2020).

[29] Fukuzawa, A., Katagiri, K., Harada, K., Masumoto, K., Chogahara, M., Kondo, N., & Okada, S. (2019). A longitudinal study of the moderating effects of social capital on the relationships between changes in human capital and ikigai among Japanese older adults. *Asian Journal of Social Psychology*, *22*(2), 172-182. https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajsp.12353 (accessed 19 Sep 2020).

[30] Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565. https://arxiv.org/pdf/1606.06565.pdf%20http://arxiv.org/abs/1606.06565.pdf (accessed 19 Sep 2020).

[31] Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. In N. Bostrom & M.M. Ćirković (Eds.), *Global catastrophic risks*. New York: Oxford University Press, 308-345. https://intelligence.org/files/AIPosNegFactor.pdf (accessed 19 Sep 2020).

[32] Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies; Oxford University Press: Oxford, UK.

[33] Soares, N. (2015). The value learning problem. Machine Intelligence Research Institute, Berkeley. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.674.6424&rep=rep1&type=pdf (accessed 19 Sep 2020).

[34] Turchin, A., & Denkenberger, D. (2019). Literature Review: What Artificial General Intelligence Safety Researchers Have Written About the Nature of Human Values. https://philpapers.org/rec/TURLRW (accessed 19 Sep 2020).

[35] Omohundro, S.M. (2007). The nature of self-improving artificial intelligence. Singularity Summit. https://selfawaresystems.files.wordpress.com/2008/01/nature_of_self_improving_ai.pdf (accessed 19 Sep 2020).

[36] Nye, B. D. & Silverman, B. G. (2012). Affordances in AI. In N. M. Seel (Ed.), Encyclopedia of the Sciences of Learning. New York, NY: Springer, 183-187. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1677&context=ese_papers (accessed 19 Sep 2020).

[37] Yampolskiy, R. V. (2014). Utility function security in artificially intelligent agents. *Journal of Experimental & Theoretical Artificial Intelligence*, *26*(3), 373-389. http://cecs.louisville.edu/ry/Utility.pdf (accessed 19 Sep 2020).

[38] Bostrom, N. (2003). Ethical issues in advanced artificial intelligence. *Science fiction and philosophy: from time travel to superintelligence*, 277-284. http://www.fhi.ox.ac.uk/wp-content/uploads/ethical-issues-in-advanced-ai.pdf (accessed 19 Sep 2020).

[39] Ziesche, S. (2019). An AI May Establish A Religion. In *Death and Anti-Death*, ed. C. Tandy, Volume 17: 309-334. Ann Arbor: Ria University Press.

[40] Tomasik, B. (2011). Risks of astronomical future suffering. *Foundational Research Institute: Berlin, Germany*. https://longtermrisk.org/files/risks-of-astronomical-future-suffering.pdf (accessed 19 Sep 2020).

[41] Bostrom, N., Dafoe, A., & Flynn, C. (2018). Public Policy and Superintelligent AI: A Vector Field Approach. *Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK*. https://nickbostrom.com/papers/aipolicy.pdf (accessed 19 Sep 2020).

[42] Ziesche, S. & Yampolskiy, R. V. (2019a). Towards AI Welfare Science and Policies. *Big Data and Cognitive Computing, 3(1):2*. https://www.mdpi.com/2504-2289/3/1/2/htm (accessed 19 Sep 2020).

[43] Aliman, N. M., Elands, P., Hürst, W., Kester, L., Thórisson, K. R., Werkhoven, P., Yampolskiy, R. & Ziesche, S. (2020). Error-Correction for AI Safety. In *International Conference on Artificial General Intelligence*, 12-22. Springer, Cham.

[44] Tomasik, B. (2014). Do video-game characters matter morally? Updated June 14, 2019. https://reducing-suffering.org/do-video-game-characters-matter-morally/ (accessed 19 Sep 2020).

[45] Ziesche, S., & Yampolskiy, R. V. (2019b). Do No Harm Policy for Minds in Other Substrates. *Journal of Evolution & Technology*, *29*(2). https://jetpress.org/v29.2/ziesche.html (accessed 19 Sep 2020).

[46] Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, *1*, 316-334. http://faculty.smcm.edu/acjamieson/s13/artificialintelligence.pdf (accessed 19 Sep 2020).