Well-being

Prioritizing ethical and responsible artificial intelligence has become a widespread goal for society. Important issues of transparency, accountability, algorithmic bias, and others are being directly addressed in the design and implementation of autonomous and intelligent systems (A/IS). While this is an encouraging trend, a key question still facing technologists, manufacturers, and policy makers alike is, what should be the specific metrics of societal success for "ethical AI" once it's being used?

For A/IS to demonstrably advance the well-being of humanity, there needs to be concise and useful indicators to measure those advancements. However, there is not a common understanding of what well-being indicators are, or which ones are available. Technologists will use best-practice metrics available even if, unbeknownst to them, said metrics are inappropriate or, worse, potentially harmful. To avoid unintended negative consequences and to increase value for users and society, clear guidance on what well-being is and how it should be measured is needed.

Common metrics of success include profit, gross domestic product (GDP), consumption levels, occupational safety, and economic growth. While important, these metrics fail to encompass the full spectrum of well-being for individuals or society. Psychological, social, and environmental factors matter. Where these factors are not given equal priority to fiscal metrics of success, technologists risk causing or contributing to negative and irreversible harms to our planet and population.

This document identifies examples of existing well-being metrics that capture such factors, allowing the benefits of A/IS to be more comprehensively evaluated. While these indicators vary in their scope and use, they expand the focus of impact to aspects of human well-being that are not currently measured in the realms of A/IS.

When properly utilized, these metrics could provide an opportunity to test and monitor A/IS for unintended negative consequences that could diminish human well-being. Conversely, these metrics could help identify where A/IS would increase human well-being, providing new routes to societal and technological innovation. By corollary, A/IS can also increase the measurement and efficiency of well-being indicators.

Well-being

This Committee, along with the IEEE P7010™ Standard Working Group, Well-being Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems, was created with the belief that A/IS should prioritize human well-being as an outcome in all system designs, using the best available and widely accepted well-being metrics as their reference point.

This document is divided into the following sections:

- An Introduction to Well-being Metrics *(What you need to know)*
- The Value of Well-being Metrics for A/IS *(Why you should care)*
- Adaptation of Well-being Metrics for A/IS *(What you can do)*

Appendix:

The following sections are included in the Appendix as separate documents to provide readers with an introduction to existing individual and societal level well-being metrics currently in use:

- The State of Well-being Metrics. This section identifies well-being metrics being used today by social scientists, international institutions, and governments to provide an overall introduction to well-being.
- The Happiness Screening Tool for Business Product Decisions. This tool is provided as an example of how well-being indicators can inform decisions.

Disclaimer: *While we have provided recommendations in this document, it should be understood these do not represent a position or the views of IEEE but the informed opinions of Committee members providing insights designed to provide expert directional guidance regarding A/IS. In no event shall IEEE or IEEE-SA Industry Connections Activity Members be liable for any errors or omissions, direct or otherwise, however caused, arising in any way out of the use of this work, regardless of whether such damage was foreseeable.*

Section 1 — An Introduction to Well-being Metrics

This section provides a brief overview of what well-being metrics are outside of the context of A/IS to provide a background for readers who may not be familiar with these areas.

Issue:

There is ample and robust science behind well-being metrics and use by international and national institutions, yet many people in the A/IS field and corporate communities are unaware that well-being metrics exist, or what entities are using them.

Background

The concept of *well-being* refers to an evaluation of the general goodness of a state or event to the individual or community as a distinct moral or legal evaluation. The term itself has been used and defined in various ways across different contexts and fields. For the purposes of this committee, well-being is defined as encompassing human satisfaction with life and the conditions of life, flourishing (eudaimonia), and positive and negative affect, following the Organization for Economic Cooperation and Development (OECD) Guidelines on Measuring Subjective Well-being (p. 12). This holistic definition of well-being encompasses individual, social, economic, and governmental circumstances as well as human rights, capabilities, environmental protection, and fair labor, as these circumstances and many others form the basis for human well-being.

Well-being metrics fall into four categories:

1. **Subjective or survey-based indicators**

- Survey-based or subjective well-being (SWB) indicators are being used by international institutions and countries to understand levels of reported well-being within a country and for aspects of citizen demographics. Examples include the European Social Survey, Bhutan's Gross National Happiness Indicators, and well-being surveys created by The UK Office for National Statistics. Survey-based or subjective metrics are also employed in the field of positive psychology and in the World Happiness Report, and the data are employed by researchers to understand the causes, consequences, and correlates of well-being as subjects see it. The findings of these researchers provide crucial and necessary guidance to policy makers, leaders, and others in making decisions regarding people's subjective sense of well-being.

Well-being

2. Objective indicators

- Objective well-being indicators have been used to understand conditions enabling well-being of countries and to measure the impact of companies. They are in used by organizations like the OECD with their Better Life Index (which also includes subjective indicators), and United Nations with their Millennium Development Goal Indicators. For business, the Global Reporting Initiative, SDG Compass, and B-Corp provide broad indicator sets.

3. Composite indicators (indices that aggregate multiple metrics)

- Aggregate metrics combine subjective and/or objective metrics to produce one measure. Examples of this are the UN's Human Development Index, the Social Progress Index, and the United Kingdom's Office of National Statistics Measures of National Well-being.

4. Social media sourced data

- Social media is source used to measure the well-being of a geographic region or demographics, based on sentiment analysis of publicly available data. Examples include the Hedonometer and the World Well-being Project.

The appendix The State of Well-being Metrics *provides a broad primer on the state of well-being metrics.*

Candidate Recommendation

A/IS policy makers and manufacturers (including academics, designers, engineers, and corporate employees) should prioritize having all their stakeholders learn about well-being metrics as potential determinants for how they create, deploy, market, and monitor their technologies. This process can be expedited by having organizations including the Global Reporting Initiative (GRI), B-Corp, and Standards Development Organizations (SDO) create certifications, guidelines, and standards that demonstrate the value of holistic, well-being-centric reporting guidelines for the A/IS public and private sectors.

Further Resources

- The IEEE P7010™ Standards Working Group, Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems has been formed with the aim of identifying well-being metrics for applicability to A/IS today and in the future. All are welcome to join the working group.
- On 11 April 2017, IEEE hosted a dinner debate at the European Parliament in Brussels to discuss how the world's top metric of value *(gross domestic product)* must move Beyond GDP to holistically measure how intelligent and autonomous systems can hinder or improve human well-being:
  - Prioritizing Human Well-being in the Age of Artificial Intelligence *(Report)*
  - Prioritizing Human Well-being in the Age of Artificial Intelligence *(Video)*

Section 2 — The Value of Well-being Metrics for A/IS

Well-being metrics, in the form of triple-bottom line benefits ("people, planet, and profit") for the corporate world, and in the form of tools to measure a population's well-being for policy makers, can provide value to A/IS technologists. Where technologists may be unaware of how systems could negatively impact human well-being, by increasing awareness of common indicators and their designed intent, they can avoid harm while increasing benefit.

In addition, a key value for the use of well-being metrics for A/IS technologists comes in the form of predictive modeling (forecasting outcomes based on data analysis and probabilities), either for unintended consequences, or as a unique means of innovation regarding metrics or areas of consideration not currently being measured today.

Issue:

Many people in the A/IS field and corporate communities are not aware of the value well-being metrics offer.

Background

While many organizations are aware of the need to incorporate sustainability measures as part of their efforts, the reality of bottom line, quarterly driven shareholder growth is a traditional metric prioritized within society at large. Where organizations exist in a larger societal ecosystem equating exponential growth with success, as mirrored by GDP or similar financial metrics, these companies will remain under pressure to deliver results that do not fully incorporate societal and environmental measures and goals along with existing financial imperatives.

Along with an increased awareness of how incorporating sustainability measures beyond compliance can benefit the positive association with an organization's brand in the public sphere, by prioritizing the increase of holistic well-being, companies are also recognizing where they can save or make money and increase innovation in the process.

For instance, where a companion robot outfitted to measure the emotion of seniors in assisted living situations might be launched with a typical "move fast and break things" technological manufacturing model, prioritizing largely fiscal metrics of success, these devices might fail in the market because of limited adoption. However, where they also factor in data aligning with

Well-being

uniform metrics measuring emotion, depression, or other factors (including life satisfaction, affect, and purpose), the device might score very high on a well-being scale comparable to the Net Promoter Score widely used today. If the device could significantly lower depression according to metrics from a trusted source like the World Health Organization, academic institutions testing early versions of systems would be more able to attain needed funding to advance an A/IS well-being study overall. While these are hypothetical scenarios, they are designed to demonstrate the value of linking A/IS design to well-being indicators where possible.

This is a key point regarding the work of this Committee — rather than focus on the negative aspects of how A/IS could harm humans, the implementation of uniform well-being metrics will help provably demonstrate how these technologies can have a positive influence on society.

The good news in regards to this subject is that thought leaders in the corporate arena have recognized this multifaceted need to utilize metrics beyond fiscal indicators. In 2013, PricewaterhouseCoopers released a report called Total Impact Approach: What CEOs Think from PricewaterhouseCoopers: (where total impact refers to a "holistic view of social, environmental, fiscal and economic dimensions") where they noted:

> *187 CEOs across the globe shared their views on the value of measuring total impact. From all industries, they explored the benefits, opportunities and challenges of a total impact approach. There's an overwhelming consensus (85% CEOs) that results from a total impact approach would be more insightful than financial analysis alone. Business leaders saw the more holistic perspective useful in not only managing their business, but also in communicating with certain stakeholders. But less than 25% of CEOs measure their total impact with the lack of availability of data or a robust framework holding them back.*

This report, along with more recent work being done by other thought-leading organizations in the public sector like the OECD in their February, 2017 Workshop, Measuring Business Impacts on People's Well-Being, demonstrates the desire for business leaders to incorporate metrics of success beyond fiscal indicators for their efforts. The B-Corporation movement has even created a new legal status for "a new type of company that uses the power of business to solve social and environmental problems." Focusing on increasing "stakeholder" value versus shareholder returns alone, forward-thinking B-Corps are building trust and defining their brands by provably aligning their efforts to holistic metrics of well-being.

From a mental health perspective, well-being is also important to business. Happy workers are more productive than employees who are not engaged in their careers. There are also fewer issues with absenteeism: people miss work less and have fewer health claims.

Well-being

Candidate Recommendation

A/IS and well-being experts should work directly with the business community to identify existing metrics or combinations of indicators that would bring the greatest value to businesses focused on the "triple bottom line" (accounting for economic, social, and environmental impacts) increase of human well-being. (Noting, however that well-being metrics should only be used with consent, respect for privacy, and with strict standards for collection and use of these data).

In addition, any stakeholders creating A/IS in the business or academic, engineering, or policy arenas are advised to review the Appendix listing well-being metrics to familiarize themselves with existing indicators already relevant to their work.

Further Resources

- PwC. Total Impact Approach: What CEOs Think.
- World Economic Forum. The Inclusive Growth and Development Report. January 16, 2017. Geneva, Switzerland: World Economic Forum.

Issue:

By leveraging existing work in computational sustainability or using existing indicators to model unintended consequences of specific systems or applications, well-being could be better understood and increased by the A/IS community and society at large.

Background

To date, there does not exist a definitive well-being metric that encompasses every aspect of individual and societal well-being that could serve as a common metric like the GDP for all A/IS manufacturers. Moreover, data may or may not exist within the context one wishes to measure or improve.

Modeling for Unintended Consequences

There is a potential for synergy when adapting well-being indicators for the use of A/IS. This potential is in avoiding unintended consequences. Two challenges to face when exploring this potential are: (1) Identifying which indicators to select to model potential unintended consequences; and, (2) Understanding how to predict unintended consequences when data are lacking or are incomplete.

Well-being

Machine-learning and other tools have the ability to map out potential consequences with greater specificity and efficiency than humans. In this way, A/IS could be utilized to map out potential consequences regarding how products, services, or systems might affect end users or stakeholders in regards to specific well-being indicators. In this way, models could be run during the design phase of a system, product, or service to predict how it could improve or potentially harm end users, analogous to human rights assessments provided by the United Nations Guiding Principles Reporting Framework.

As the exchange of A/IS related data regarding an individual (via personalized algorithms, in conjunction with affective sensors measuring and influencing emotion, etc.) and society (large data sets representing aggregate individual subjective and objective data) is widely available via establishing tracking methodologies, this data should be classified to match existing well-being indicators so devices or systems can be provably aligned to the increase of human well-being (satisfaction with life and the conditions of life, positive affect, and eudaimonic well-being).

As an example, today popular robots like Pepper are equipped to share data regarding their usage and interaction with humans to the cloud. This allows almost instantaneous innovation, as once an action is validated as useful for one Pepper robot, all other units (and ostensibly their owners) benefit as well. As long as this data exchange happens via pre-determined consent with their owners, this "innovation in real-time" model can be emulated for the large-scale aggregation of information relating to existing well-being metrics.

A crucial distinction between well-being metrics and potential interventions in their use is that a well-being metric does not dictate an intervention, but points the way for developing an intervention that will push a metric in a positive direction. For example, a team seeking to increase the well-being of people using wheelchairs found that when provided the opportunity to use a smart wheelchair, some users were delighted with the opportunity for more mobility, while others felt it would decrease their opportunities for social contact and lead to an overall decrease in their well-being. The point being that even increased well-being due to a smart wheelchair does not mean that this wheelchair should automatically be adopted. Well-being is only one value in the mix for adoption, where other values to consider would be human rights, respect, privacy, justice, freedom, culture, etc.

Computational Sustainability

Computational sustainability is an area of study within the A/IS community that demonstrates that the A/IS community is already showing interest in well-being even when not using this term, as the concept of sustainability encompasses aspects of well-being.

Computational sustainability directly relates to the use of these technologies to increase social good in ways that could be uniquely tied to existing well-being metrics. As defined by The Institute of Computational Sustainability, the field is designed to provide "computational models for a sustainable environment, economy, and society" and their project summary notes that:

Well-being

Humanity's use of Earth's resources is threatening our planet and the livelihood of future generations. Computing and information science can — and should — play a key role in increasing the efficiency and effectiveness in the way we manage and allocate our natural resources. We propose an expedition in Computational Sustainability, encompassing computational and mathematical methods for a sustainable environment, economy, and society.

AAAI, (the Association for the Advancement of Artificial Intelligence) the world's largest global body dedicated to the advancement of artificial intelligence had a special track on computational sustainability at their 2017 conference. The description of the track provides helpful specifics demonstrating the direct alignment between the work of this Committee and the A/IS community at large:

> *This special track invites research papers on novel concepts, models, algorithms, and systems that address problems in computational sustainability. We are looking for a broad range of papers ranging from formal analysis to applied research. Examples include papers explaining how the research addresses specific computational problems, opportunities, or issues underlying sustainability challenges and papers describing a sustainability challenge or application that can be tackled using AI methods. Papers proposing general challenges and data sets for computational sustainability are also welcome. All AI topics that can address computational sustainability issues are appropriate, including machine learning, optimization, vision, and robotics, and others. Sustainability domains include natural resources, climate, and the environment (for example, climate change, atmosphere, water, oceans, forest, land, soil, biodiversity, species), economics and human behavior (for example, human well-being, poverty, infectious diseases, over-population, resource harvesting), energy (for example, renewable energy, smart grid, material discovery for fuel cell technology) and human-built systems (for example, transportation systems, cities, buildings, data centers, food systems, agriculture).*

Candidate Recommendations

- Work with influencers and decision-makers in the computational sustainability field to cross-pollinate efforts of computational sustainability in the A/IS field and the well-being communities to expedite efforts to identify, align, and advance robust and uniform indicators into current models that prioritize and increase human well-being. Develop cross-pollination between the computational sustainability and well-being professionals to ensure integration of well-being into computational sustainability, and vice-versa.
- Explore successful programs like LEED Building Design Standards, ISO 2600 Corporate Responsibility, ISO 37101 Sustainable Development Standards, and others to determine what new standards or certification models along these lines approach would be valuable and operationalizable for A/IS.

Well-being

Further Resources

- Gomes, C. P. "Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society" in *The Bridge: Linking Engineering and Society.* Washington, DC: National Academy of Engineering of the National Academies, 2009.
- Meadows, D. H., D. L. Meadows, J. Randers, and W. W. Behrens, III. The Limits to Growth. New York: Universe Books, 1972. Reissued in 2004 by Chelsea Green Publishing & Earthscan.
- LEED Building Design Standards program.
- ISO 2600, Guidance on Social Responsibility.
- ISO 37101, Sustainable Development in Communities

Issue:

Well-being indicators provide an opportunity for modeling scenarios and impacts that could improve the ability of A/IS to frame specific societal benefits for their use.

Background:

There is a lack of easily available or widely recognized scenarios along these lines.

Candidate Recommendation

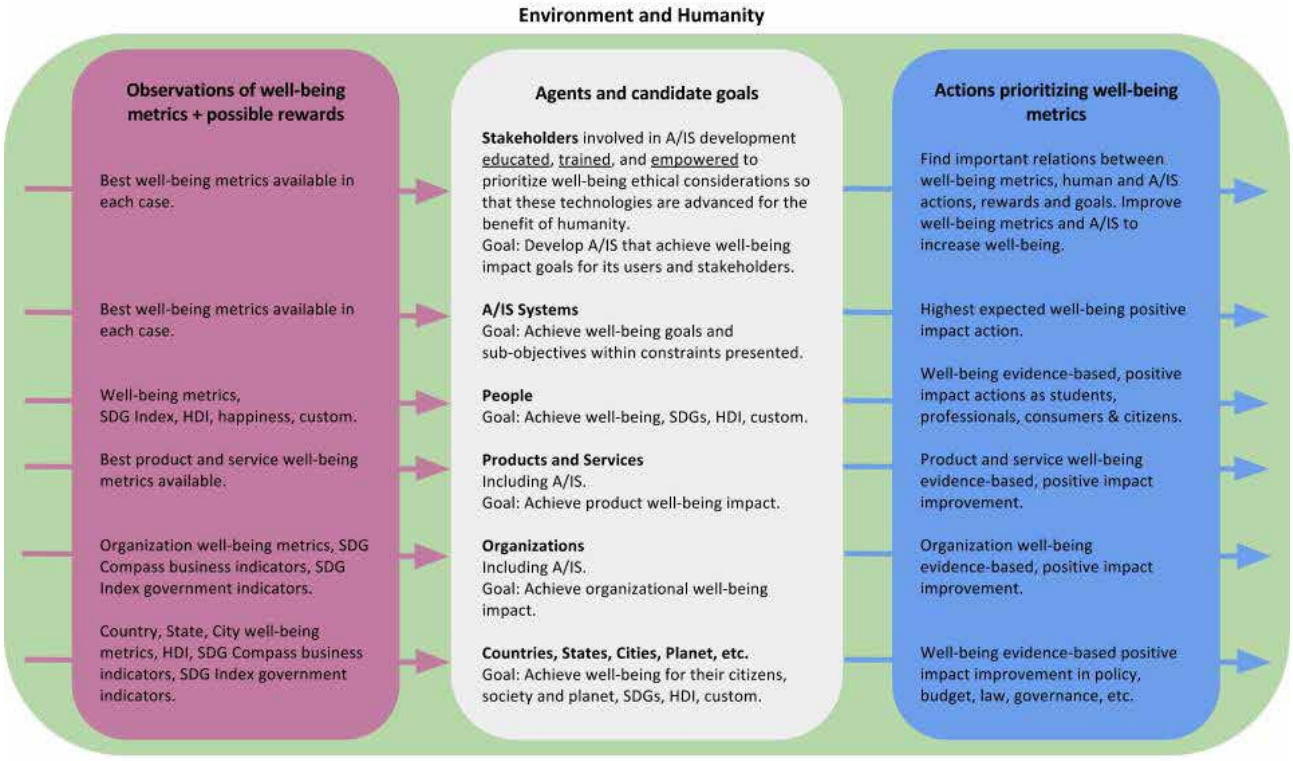
Rigorously created well-being assessments could be utilized as a public "scoreboard," or statement of intent, that would provide innovation opportunities for technologists as well as a form of public accountability for human sustainability.

Further Resources

The following schema and well-being assessment tool provide an initial attempt to visualize how A/IS technologists can utilize well-being metrics in their work. By modeling the potential positive or negative impacts of technologies across a full spectrum of financial, environmental, and social impacts (e.g., a "triple bottom line" well-being indicator model) A/IS technologists can better avoid negative unintended consequences for human well-being, while increasing innovation and positive human well-being for their work.

Well-being

Schema of A/IS and the Stakeholders Involved in Their Development



Schema of A/IS systems and the stakeholders involved in their development, **adapting and operationalizing well-being metrics for ethical A/IS** in a world model.

The schema represents a model of the world where the stakeholders (designers, engineers, technologists, researchers, managers, users, etc.) involved in A/IS development adapt and operationalize well-being metrics for ethical A/IS. Stakeholders can visualize important entities in the world as agents with different goals that receive observations and possible rewards from the environment and make actions that could have positive and negative impacts to the well-being of different agents.

This schema could help to assess, in different cases, the well-being metrics that the A/IS should take into account and the well-being metrics of the impacts that A/IS actions could and can cause, related to important elements in the world like: people, products, organizations, climate, countries, etc. An applied case of this schema could be seen in the following well-being impact assessment.

Well-being

Well-being Impact Assessment

Here is a concept for simple A/IS well-being impact assessment, based on Maslow's Hierarchy of Need (where the Hierarchy would be considered an accredited and contextually appropriate metric of use). Given that a working definition of well-being including both individual and societal key performance indicators (KPIs) is still being developed, this metric is general and used for illustrative purposes only.

Please also note that this is a purely conceptual framework used as a directional teaching tool for readers. It doesn't yet include an evaluative component or reflect the holistic nature of well-being at this time like The Happiness Screening Tool (based on the government of Bhutan's Policy Screening Tool) provided in the Appendix. It should be noted that any impact assessment created by A/IS and well-being experts working together identify best-in-class (existing) metrics within specific contexts of use.

	Individual Direct	Individual Indirect	Environment Direct	Individual Indirect	Social Direct	Social Indirect
Basic Needs						
Safety						
Belonging						
Esteem						
Self-Actualization						
Overall Impact						

Indicators:

nil impact = 0 *negative impact = –* *positive impact = +* *unknown impact = ?*

Well-being

The following examples are provided to demonstrate specific A/IS applications within this framework and include: a retail kiosk robot, a small factory arm, a mental health chatbot, and a companion robot. The goal of these diagrams is to provide a sample of how the work of matching established well-being metrics to A/IS work could progress.

Retail Kiosk Robot	Individual Direct	Individual Indirect	Environment Direct	Individual Indirect	Social Direct	Social Indirect
Basic Needs	0	0	0	–	+	?
Safety	?	?	?	–	+	?
Belonging	+	?	+	?	+	?
Esteem	+	?	0	0	+	?
Self-Actualization	?	?	0	0	?	?
Overall Impact	Mild +	Unknown	Very Mild +	Mild –	Strong +	Unknown

Using this tool, the retail kiosk robot scores are mildly beneficial in the category of Individual Direct (i.e., reduced barriers to goal attainment) and Environmental Direct (i.e., use of resources), while strongly beneficial in Social Direct (i.e., better access to mental health support), but mildly unbeneficial in Environment Indirect (i.e., carbon footprint), and unknown in Social Indirect (i.e., job loss) categories. The robot is "helpful and kind," but of limited utility or interaction value. Another example of a negative impact on well-being is gendering, racial identification, or physical attributes of kiosk robots (such as a slim, youthful appearing, Caucasian, female), leading to harmful stereotyping.

Well-being

Small Factory Arm	Individual Direct	Individual Indirect	Environment Direct	Individual Indirect	Social Direct	Social Indirect
Basic Needs	+	+	0	–	+	+
Safety	?	?	?	–	?	+
Belonging	–	–	0	0	0	0
Esteem	–	–	0	0	0	0
Self-Actualization	0	0	0	0	0	0
Overall Impact	Mild –	Mild –	Nil	Mild –	Mild +	Mild +

The tool indicates that robots need to be assessed more thoroughly on their safe operations to better answer impact assessment, and that this is also a robot with very limited interaction with people. But the diagram shows how the arm could have a potentially negative impact on self-worth and belonging, but a positive impact on basic needs both for individuals and society.

Mental Health Chatbot	Individual Direct	Individual Indirect	Environment Direct	Individual Indirect	Social Direct	Social Indirect
Basic Needs	0	0	0	0	0	0
Safety	+	0	0	0	?	+
Belonging	+	?	0	0	?	–
Esteem	+	?	0	0	?	–
Self-Actualization	?	0	0	0	0	0
Overall Impact	Strong +	Unknown	Nil	Nil	Unknown	Mild –

Well-being

There is evidence that a mental health aide chatbot could improve individual self esteem and ultimately reduce self harm, but there is little evidence supporting claims that this would improve society directly or indirectly. The reliance on artificial support may have a net negative impact on society. However, this would need to be determined by the A/IS and well-being experts applying this methodology once created in a robust and rigorous manner.

Companion Robot like Paro	Individual Direct	Individual Indirect	Environment Direct	Individual Indirect	Social Direct	Social Indirect
Basic Needs	0	0	0	-	0	0
Safety	+	?	0	0	0	0
Belonging	+	?	0	0	?	-
Esteem	+	?	0	0	?	-
Self-Actualization	?	0	0	0	0	0
Overall Impact						

For a small resource cost, a companion robot can provide significant psychological assistance. On the one hand, this makes society more caring, but on the other hand reliance on artificial companionship shows a lack of social resources in this area. A potential negative impact is development of reliance on companionship and negative impact on people who lose access to companion robot.

The Happiness Project Screening Tool for Business provided in the Appendix could also augment this if a product shows a low or negative score in the areas of well-being. Another set of metrics that could be used in a more detailed schema are the Kingdom of Bhutan's nine domains of well-being: psychological well-being, health, community vitality, living standards, governance, environment diversity, culture, education, and time use.

Whatever established well-being metrics that may be utilized for such a methodology, it is critical for A/IS technologists and well-being experts to work in unison to create assessment tools using best in class data, indicators, and practices in their potential analysis and use.

Section 3 — Adaptation of Well-being Metrics for A/IS

This section focuses on areas of immediate attention for A/IS technologists to be aware of regarding well-being metrics in an effort to aid their work and avoid negative unintended consequences.

Issue:

How can creators of A/IS incorporate measures of well-being into their systems?

Background

Just as undirected A/IS can lead to negative outcomes, A/IS directed only to specific ends without considering human well-being can lead to negative side effects. Without practical ways of incorporating widely shared ways of measuring and promoting well-being metrics and expected well-being outcomes available to designers, A/IS will likely lack beneficence.

Once well-being metrics are widely recognized as a directional requirement for society, conceptually, one would like such measures to be supported by the engines of change and leverage within society. A/IS will be an increasing portion of such engines. How might designers architect systems to include such measures as considerations while executing their primary objectives? How will these measures be adapted as we learn more?

Existing metrics of well-being could be formulated into a sub-objective of the A/IS. In order to operate with respect to such sub-objectives, it is instrumental to evaluate the consequences of the A/IS's actions. As practical systems are bounded and can predict over only limited horizons, it may be necessary to supplement these evaluations with both biases toward virtues and deontological guidelines or soft constraints as lesser supplemental components, informed by the well-being metrics and their precursors or constituents.

As these well-being sub-objectives will be only a subset of the intended goals of the system, the architecture will need to balance multiple objectives. Each of these sub-objectives may be expressed as a goal, or as a set of rules, or as a set of values, or as a set of preferences, and those can be combined as well, using established methodologies from intelligent systems engineering.

For example, people, organizations, and A/IS, collaborating together, could understand the well-being impacts and objectives of products, services, organizations, and A/IS within the context of the well-being of communities, cities, countries, and the planet using the SDG Index

Well-being

and Dashboards, the SDG Compass Inventory of Business Indicators and other metrics. This collaboration of people, organizations and A/IS could make decisions and take actions with high expected utility to well-being objectives and goals such as those stated in the Sustainable Development Goals and similar institutions. This collaboration could lead to a more humane, organizational, and computational sustainability for individuals, all of society, and the planet.

International organizations, lawmakers, and policy experts can specify core values and/or sub-objectives as rules for the benefit of society utilizing well-being metrics as a starting point and these can be pluggable and hierarchical by jurisdiction. Similarly, industry organizations would be able to specialize norms and industry self-regulation (e.g., any automated flight attendants should prevent onboard smoking and sit down during takeoff) as a layer.

System designers should ensure situational awareness as well as prediction of the consequences of their actions based on some world model. They could also layer in their own sub-objectives and make the system's values explicit.

Resellers, service organizations, or owners that have particular primary goals for their systems would still be able to specify primary goals for the system (e.g., mowing lawns, doing taxes, etc.), and those would be alongside the other deeper-down subgoals and values as well for societal benefit, public safety, etc., directly relating to established well-being metrics.

End users would have the opportunity to layer on their own preferences in these systems, and would also be able to get an explanation and inventory of the types of objectives or value systems the A/IS holds relating to established well-being metrics, including what permissioning is required for modifying or removing them.

Candidate Recommendation

Formation of a working group to develop a blueprint for the fluid and evolving (institutional learning) operationalization of A/IS well-being indicators for the various stakeholders (e.g., technicians, coders, and system designers), international well-being oriented organizations, lawmakers, and policy experts, industry organizations, retailers, resellers, service organizations and owners, and end users.

Candidate Recommendation

Creation of technical standards for representing dimensions, metrics, and evaluation guidelines for well-being metrics and their precursors and constituents within A/IS. This would include ontologies for representing requirements as well as a testing framework for validating adherence to well-being metrics and ethical principles. (For more information, please see IEEE P7010™ Standards Working Group mentioned above).

Well-being

Further Resources

- Calvo, R. A., and D. Peters. Positive Computing: Technology for Well-Being and Human Potential. Cambridge MA: MIT Press, 2014
- Collette Y., and P. Slarry. Multiobjective Optimization: Principles and Case Studies (Decision Engineering Series). Berlin, Germany: Springer, 2004. doi: 10.1007/978-3-662-08883-8.
- Greene, J. et al. "Embedding Ethical Principles in Collective Decision Support Systems," in: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence,* 4147–4151. Palo Alto, CA: AAAI Press, 2016.
- Li, L. et al. "An Ontology of Preference-Based Multiobjective Evolutionary Algorithms," 2016. CoRR abs/1609.08082.
- A. FT Winfield, C. Blum, and W. Liu. "Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection," in *Advances in Autonomous Robotics Systems*. Springer, 2014, pp. 85–96.
- Gershman, S. J., E. J. Horvitz, and J. B. Tenenbaum. "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines." *Science* 349, no. 6245 (2015): 273–278.
- PositiveSocialImpact: Empowering people, organizations and planet with information and knowledge to make a positive impact to sustainable development.

Issue:

A/IS technologies designed to replicate human tasks, behavior, or emotion have the potential to either increase or decrease well-being.

Background

A/IS are already being executed in ways that could dramatically increase human well-being or, possibly, have an undue coercive effect on humans.

A/IS technologies present great opportunity for positive change in every aspect of society. However, sophisticated manipulative technologies utilizing A/IS can also restrict the fundamental freedom of human choice, and are able to manipulate humans who consume customized content without recognizing the extent of manipulation. Software platforms are moving from targeting content to much more powerful and potentially harmful "persuasive computing." A/IS with sophisticated manipulation technologies (so-called "big nudging") will be able to guide individuals through entire courses of action, whether it be a complex work process, consumption of free content, or political persuasion.

There is also a related concern that big nudging can be done without anyone realizing harm is occurring. With deep learning methods,

Well-being

technologies may not be understood, much less contemplated. This begs the age-old question: just because one can do something, does that mean one should? Hence, there is a need to understand A/IS well-being related processes and impacts further, and to devise ways to protect people from harm and secure well-being in the furtherance of A/IS.

A/IS may also deceive and harm humans by posing as humans. With the increased ability of artificial systems to meet the Turing test (an intelligence test for a computer that allows a human to distinguish human from artificial intelligence), there is a significant risk that unscrupulous operators will abuse the technology for unethical commercial, or outright criminal, purposes. The widespread manipulation of humans by A/IS and loss of human free agency, autonomy, and other aspects of human flourishing, is by definition a reduction in human well-being. Without taking action to prevent it, it is highly conceivable that A/IS will be used to deceive humans by pretending to be another human being in a plethora of situations or via multiple mediums.

Without laws preventing A/IS from simulating humans for purposes like deception and coercion, and enforcing A/IS to clearly identify as such, mistaken identity could also reasonably be expected.

Candidate Recommendation

To avoid potential negative unintended consequences, A/IS manufacturers, and society in general, should prioritize the analysis and implementation of practices and policy that secures or increases human well-being, including:

- Well-being metrics to guide the development and implementation of A/IS should increase human well-being, defined subjectively in terms of cognitive, affective, and eudaimonic domains, and objectively in terms of conditions enabling well-being.
- While individuals may enjoy the ability of A/IS to simulate humans in situations where they are pure entertainment, explicit permission and consent by users in the use of these systems is recommended, and the well-being impacts on users should be monitored, researched, and considered by the A/IS community in an effort to provide services and goods that improve well-being. As part of this, it is important to include multiple stakeholders, including minorities, the marginalized, and those often without power or a voice.
- The implications of A/IS on human well-being are important issues to research and understand. A literature review to determine the status of academic research on the issue of A/IS impacts on human well-being needs to be conducted and aggregated in a centralized repository for the A/IS community.

Well-being

Further Resources

- Helbing, D. et al. "Will Democracy Survive Big Data and Artificial Intelligence?" *Scientific American*, February 25, 2017.
- Schenker, J. L. "Can We Balance Human Ethics with Artificial Intelligence?" *Techonomy,* January 23, 2017.
- Bulman, M. "EU to Vote on Declaring Robots To Be 'Electronic Persons." *Independent,* January 14, 2017.
- Nevejan, N. for the European Parliament. "European Civil Law Rules in Robotics." October 2016.
- "The AI That Pretends To Be Human," *LessWrong* blog post, February 2, 2016.
- Chan, C. "Monkeys Grieve When Their Robot Friend Dies." *Gizmodo*, January 11, 2017.

Issue:

Human rights law is sometimes conflated with human well-being, leading to a concern that a focus on human well-being will lead to a situation that minimizes the protection of inalienable human rights, or lowers the standard of existing legal human rights guidelines for non-state actors.

Background

International human rights law has been firmly established for decades and the protection of human rights must be an end result in itself. Some countries or regimes have highlighted the use or increase of certain "well-being" measures as justification to violate human rights, as happens in countries that conduct ethnic cleansing or mistreat refugees or immigrants who are portrayed as threatening a nation's culture or economic structure.

While the use of well-being metrics to justify human rights violations is an unconscionable perversion of the nature of any well-being metric, these same practices happen today in relation to the GDP. For instance, today, according to the International Labor Organization (ILO) approximately 21 million people are victims of forced labor (slavery) representing between 9% to 56% of various countries current GDP income. These clear human rights violations, from sex trafficking and child armies, to indentured farming or manufacturing labor, increase a country's GDP.

Well-being metrics and mechanisms should also take into consideration, and happen in conjunction with, independent assessments on respect and international obligations to promote, protect, and fulfill a full spectrum of human rights. For example, the use of the goal of well-being in the context of repairing and enhancing humans, predictive policing, or autonomous weapons systems to protect the public may have negative impacts on the rights of individuals or groups. Moreover, the development and delivery of A/IS should adopt a human rights approach to technology, including, but not limited to, the

Well-being

UN Guiding Principles on Human Rights (also known as the Ruggie principles).

To avoid issues of conflation and confusion, it is critical to note the following: human rights involves adhering to the firmly established application of international human rights law. Well-being metrics are designed to measure the efficacy of the implementation of methodologies and policy related to individual and societal flourishing.

Well-being as a value is also distinct from justice, responsibility, and freedom. But A/IS technologies can be narrowly conceived from an ethical standpoint and still be legal and safe in their usage following existing practices, but not contribute to human well-being. In this regard, well-being considerations do not displace other issues of human rights or ethical methodologies, but rather complement them.

Candidate Recommendation

Human rights and human well-being should not be held as trade-offs, with one to be prioritized over the other. In this regard, well-being metrics can be complementary to the goals of human rights, but cannot and should not be used as a proxy for human rights or any existing law.

Further Resources

- Project Include - The site features an open source manual for creating diversity in tech and highlights three key points for creating change: inclusion, comprehensiveness, and accountability.
- OpenDiversityOrg initiative from Double Union and Project Include have an action document with a lot of recommendations.
- "The Diversity Debt" by Susan Wu at Project Include is a compelling example of converting a problem into innovation language.

Issue:

A/IS represents opportunities for stewardship and restoration of natural systems and securing access to nature for humans, but could be used instead to distract attention and divert innovation until the planetary ecological condition is beyond repair.

Background

Human well-being, the existence of many other species, as well as economic and social systems, draw from and depend upon healthy ecological systems and a healthy local and planetary environment. Research using geo-data finds that human well-being is enhanced through access to nature. Many bank on technology to answer the threats of climate change, water scarcity, soil degradation, species extinction, deforestation, deterioration of biodiversity, and destruction of ecosystems that threaten humankind and other life forms.

Well-being

While technology may be the answer for some of these threats, it is unclear whether benefits extend beyond those from high socio-economic class to the majority of people, particularly the middle class and working poor, as well as those suffering from abject poverty, fleeing disaster zones or otherwise lacking the resources to meet their needs. For example, in cities in China where air pollution is so prevalent that the air is unhealthy, a few schools have covered "outdoor" fields with domes full of purified air while most children must risk their lungs when playing outside, or play indoors. Moreover, it is well-understood that ecological crises, such as sea level rise and fisheries depletion, will not only negatively impact business interests, but it will have a significantly more devastating impact on the poor and developing nations than the wealthy and developed nations.

Candidate Recommendation

Well-being metrics employed for A/IS should include measures for ecological/environmental sustainability that point the direction toward stewardship and restoration of natural systems and ensure equitable environmental justice.

Candidate Recommendation

Convene a committee to issue findings on the modalities and potentials already identified in which A/IS makes progress toward stewardship and restoration of natural systems; trends in the A/IS field that represent threats to and opportunities for ecological sustainability and environmental justice; and areas for suggested future innovation and implementation.

Further Resources

- Newton, J. "Well-being and the Natural Environment: An Overview of the Evidence." August 20, 2007.
- Dasgupta, P. Human Well-Being and the Natural Environment. Oxford, U.K.: Oxford University Press, 2001.
- Haines-Young, R., and M. Potschin. "The Links Between Biodiversity, Ecosystem Services and Human Well-Being," in *Ecosystem Ecology: A New Synthesis*, edited by D. Raffaelli, and C. Frid. Cambridge, U.K.: Cambridge University Press, 2010.
- Hart, S. *Capitalism at the Crossroads: Next Generation Business Strategies for a Post-Crisis World.* Upper Saddle River, NJ: Pearson Education, 2010.
- United Nations Department of Economic and Social Affairs. "Call for New Technologies to Avoid Ecological Destruction." Geneva, Switzerland, July 5, 2011.
- Pope Francis. Encyclical Letter Laudato Si, On the Care of Our Common Home. May 24, 2015.

Well-being

Issue:

The well-being impacts of A/IS applied to human genomes are not well understood.

Background

As A/IS are increasingly used to interpret the health significance of our genomics data ("deep genomics") and to contribute to the subsequent engineering and editing of our genomes, important ethical and governance questions are in the background that provide an opportunity to utilize well-being metrics to ensure the beneficial development of genomic research as it relates to A/IS.

Imagine this scenario:

6 A.M., Washington, DC — Erika wakes up and quickly checks her "digital DNA avatar," a digital version of her genetic blueprint as it evolves day by day. The avatar knows a lot about her as it constantly monitors the interactions between her genes, analyzes her bodily fluids and diet, as well as integrates data about the air quality around her. Her avatar proposes a few advices about food choices and exercise patterns. Everything seems in check, nothing to be worried about. For now.

A first overarching reflection concerns the relationship between well-being and an increasing ability to understand and engineer our genomes: How do in-depth and personalized understanding of how our genomes function and evolve relate to the notion of well-being as measured traditionally and/or according to well-being measures? When does a reductionist interpretation of the health significance of our genomics data threaten our well-being?

Other significant questions include:

- How accurate will the predictive health data coming from the convergence of A/IS and genomics be?
- How will these health predictions be used, and who will have access to them?
- Do pharmaceutical and insurance companies have the right to use and profit from your health data predictions/modeling without giving you any benefits back in return?
- Would it threaten your self-worth if those handling your health data know a lot of biological details about your body?
- Is it ethical for a prospective employer to ask how your health will look like in the next decade?

Answers to these questions are not easy to capture, but their impact on well-being within society is profound.

The convergence of genomics technologies and A/IS creates new opportunities to define our identity and well-being within a simple narrative in which our genes have the power to tell us who

Well-being

and how well we are. As A/IS are increasingly used to interpret the health significance of our genomics data ("deep genomics") and to contribute to the subsequent engineering/editing of our genomes, we should consider important ethical and governance questions.

There is an urgent need to concurrently discuss how the convergence of A/IS and genomic data interpretation will challenge the purpose and content of relevant legislation that preserve well-being, such as, for the United States, the Health Insurance Portability and Accountability Act (HIPAA) and the Genetic Information Non-Discrimination Act (GINA). Finding the right balance of protection and regulation in using A/IS to interpret the health significance of genomics data will be important. Too much regulation could endanger precision medicine initiatives in some countries, while others would be leading the bio-race. Too little regulation could leave citizens vulnerable to different forms of threats to their well-being.

Candidate Recommendation

A working committee should be convened gathering those at the sharp end of genomics, A/IS, ethics, and governance to start a conversation with different communities to better understand the impact on well-being of the use of A/IS to interpret (and engineer) genomics data.

Candidate Recommendation

Relevant expert and legislative committees should commission a study on the impact on well-being of deep genomics, meaning at the convergence of genomics and A/IS. Such a study is recommended to encompass diverse fields of expertise in philosophy, sociology, ethics, biosafety, biosecurity, and genomics governance. Recommendations from the study should draft proposals to frame debates in legislatures and help lawmakers start developing appropriate legislation to govern A/IS applied to genomes for the well-being of society.