Well-being

Prioritizing ethical and responsible artificial intelligence has become a widespread goal for society. Important issues of transparency, accountability, algorithmic bias, and value systems 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 policymakers alike is how to assess, understand, measure, monitor, safeguard, and improve the well-being impacts of A/IS on humans. Finding the answer to this question is further complicated when A/IS are within a holistic and interconnected framework of well-being in which individual well-being is inseparable from societal, economic, and environmental systems.

For A/IS to demonstrably advance well-being, we need consistent and multidimensional indicators that are easily implementable by the developers, engineers, and designers who are building our future. This chapter is intended for such developers, engineers, and designers— referred to in this chapter as "A/IS creators". Those affected by A/IS are referred to as "A/IS stakeholders".

A/IS technologies affect human agency, identity, emotion, and ecological systems in new and profound ways. Traditional metrics of success are not equipped to ensure A/IS creators can avoid unintended consequences or benefit from unexpected innovation in the algorithmic age. A/IS creators need expanded ways to evaluate the impact of their products, services, or systems on human well-being. These evaluations must also be done with an understanding that human well-being is deeply linked to the well-being of society, economies, and ecosystems.

Today, A/IS creators largely measure success using metrics including profit, gross domestic product (GDP), consumption levels, and occupational safety. While important, these metrics fail to encompass the full spectrum of well-being impacts on individuals and society, such as psychological, social, and environmental factors. Where the priority given to these factors is not equal to that given to fiscal metrics of success, A/IS creators risk causing or contributing to negative and irreversible harms to our people and our planet.

When A/IS creators are not aware that well-being indicators, in addition to traditional metrics, can provide guidance for their work, they are also missing out on innovation that can increase well-being and societal value. For instance, while it is commonly recognized that autonomous vehicles will save lives when safely deployed, a topic of less frequent discussion is how self-

Well-being

driving cars also have the potential to help the environment by reducing greenhouse gas emissions and increasing green space. Autonomous vehicles can also positively impact well-being by increasing work-life balance and enhancing the quality of time spent during commutes.

Unless A/IS creators are made aware of the existence of alternative measures of progress, the value they provide, and the way they can be incorporated into A/IS work, technology and society will continue to rely upon traditional metrics of success. In an era where innovation is defined by holistic prosperity, alternative measures are needed more now than ever before. The 2009 Report by the Commission on the Measurement of Economic Performance and Social Progress which contributed substantially to the worldwide movement of governments using wider measures of well-being, states, "What we measure affects what we do; and if our measurements are flawed, decisions may be distorted."

We believe that A/IS creators can profoundly increase human and environmental flourishing by prioritizing well-being metrics as an outcome in all A/IS system designs—now and for the future. *The primary intended audience for this chapter is A/IS creators who are unfamiliar with the term "well-being" as it is used in the field of positive psychology and well-being studies. Our initial goal is to provide a broad introduction to qualitative and quantitative metrics and applications of well-being to educate and inspire A/IS creators. We do not prioritize or advocate for any specific indicator or methodology. For further elaboration on the definition of well-being, please see the first Issue listed in Section 1.*

This chapter is divided into two main sections:

- The Value of Well-being Metrics for A/IS Creators
- Implementing Well-being Metrics for A/IS Creators

The following resources are available online to provide readers with an introduction to existing well-being metrics and tools currently in use:

- The State of Well-being Metrics
- The Happiness Screening Tool for Business Product Decisions
- Additional Resources: Standards Development Models and Frameworks

Well-being

Section 1—The Value of Well-being Metrics for A/IS Creators

Well-being metrics provide a broader perspective for A/IS creators than they normally might be familiar with in evaluating their products. This broader perspective unlocks greater opportunities to assure a positive impact of A/IS on human well-being, while minimizing the risk of unintended negative outcomes. This section defines well-being, discusses the value of well-being metrics to A/IS creators, and notes how similar frameworks like sustainability and human rights can be complemented by incorporating well-being metrics.

Definition of Well-being

For the purposes of *Ethically Aligned Design*, the term "well-being" refers to an evaluation of the general quality of life of an individual and the state of external circumstances. The conception of well-being encompasses the full spectrum of personal, social, and environmental factors that enhance human life and on which human life depend. The concept of well-being shall be considered distinct from moral or legal evaluation.

Issue: There is ample and robust science behind well-being metrics and their use by international and national institutions. However, A/IS creators are often unaware that well-being metrics exist, or that they can be used to plan, develop, and evaluate technology.

Background

The concept of well-being refers to an evaluation of the general goodness of the state of an individual or community and is distinct from moral or legal evaluation. A well-being evaluation takes into account major aspects of a person's life, such as their happiness, success in their goals, and their overall positive functioning in their environment. There is now a thriving area of scientific research into the psychological, social, behavioral, economic, and environmental determinants of human well-being.

Well-being

The term "well-being" is defined and used in various ways across different contexts and fields. For example: economists identifying economic welfare with levels of consumption and economic vitality, psychologists highlighting subjective experience, and sociologists emphasizing living, labor, political, social, and environmental conditions. We do not take a stand on any specific measure of well-being. The metrics listed below are an incomplete list and provided as a starting point for further inquiry. Among these are subjective well-being indicators, measures of quality of life, social progress and capabilities, and many more.

There is now sufficient consensus among scientists that well-being can be reliably measured. Well-being measures differ in the number and the intricacy of indicators they employ. Short questionnaires of life satisfaction have emerged as particularly popular, although they do not reflect all aspects of well-being. While recognizing a scope for differences across well-being indicators, we note that the richest conception of well-being encompasses the full spectrum of personal, social, and environmental goods that enhance human life.

We encourage A/IS creators to consider the wide range of available indicators and select those most relevant and revealing for particular stages of the A/IS technology's life cycle and the particular context for the technology's use and evaluation. That is, measures of well-being that may be well-suited to wealthy, industrialized nations may be less applicable in low- and middle-income countries, and vice versa.

Among the most important and recognized aspects of well-being are (in alphabetical order):

- Community: Belonging, Crime & Safety, Discrimination & Inclusion, Participation, Social Support

- Culture: Identity, Values

- Economy: Economic Policy, Equality & Environment, Innovation, Jobs, Sustainable Natural Resources & Consumption & Production, Standard of Living

- Education: Formal Education, Lifelong Learning, Teacher Training

- Environment: Air, Biodiversity, Climate Change, Soil, Water

- Government: Confidence, Engagement, Human Rights, Institutions

- Human Settlements: Energy, Food, Housing, Information & Communication Technology, Transportation

- Physical Health: Health Status, Risk Factors, Service Coverage

- Psychological Health: Affect (feelings), Flourishing, Mental Illness & Health, Satisfaction with Life

- Work: Governance, Time Balance, Workplace Environment

Well-being

In an effort to provide a basic orientation to well-being metrics, information about well-being indicators can be segmented into four categories:

1. Subjective or survey-based indicators

Survey-based well-being indicators, subjective well-being (SWB) indicators, and multidimensional measurements of aspects of well-being, are being used by national institutions, international institutions, and governments to better understand levels of psychological well-being within countries and aspects of a country's population. These indicators are also being used to understand people's satisfaction in specific domains of life. Examples of surveys that include survey-based well-being indicators and SWB indicators include the European Social Survey, Bhutan's Gross National Happiness Indicators, well-being surveys created by The UK Office for National Statistics, and many more.

Survey-based metrics are also employed in the field of positive psychology and in the World Happiness Report. The data are employed by researchers to understand the causes, consequences, and correlates of well-being. Data gathered from surveys tend to address concerns, such as day-to-day experience, overall satisfaction with life, and perceived flourishing. The findings of these researchers provide crucial and necessary guidance because they often diverge from and complement the understanding of traditional conditions, such as economic growth.

2. Objective indicators

Objective indicators of quality of life have typically incorporated areas such as income, consumption, health, education, crime, housing, etc. These indicators have been used to understand conditions that support the well-being of countries and populations, and to measure the societal and environmental impact of companies. They are in use by organizations like the OECD with their Better Life Index, which also includes survey-based well-being indicators and SWB indicators, and the United Nations with their Sustainable Development Goals Indicators (formerly the Millennium Development Goals). 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 reflecting both objective aspects of quality of life and people's subjective evaluation of these. 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. Some subjective and objective indicators are also composite indicators, such as Bhutan's Gross National Happiness Index and the OECD's Better Life Index.

4. Social media sourced data

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

Well-being

Recommendation

A/IS creators should prioritize learning about well-being concepts, scientific learnings, research findings, and well-being metrics as potential determinants for how they create, deploy, market, and monitor their technologies, and ensuring their stakeholders learn the same. This process can be expedited if Standards Development Organizations (SDOs), such as the IEEE Standards Association, or other institutions such as the Global Reporting Initiative (GRI) or B-Corp, create certifications, guidelines, and standards that for the use of holistic, well-being metrics for A/IS in the public and private sectors.

Further Resources

- The IEEE P7010™ Standards Project for Well-being Metric for Autonomous/Intelligent Systems, was 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)

Issue: Increased awareness and application of well-being metrics by A/IS creators can create greater value, safety, and relevance to corporate communities and other organizations in the algorithmic age.

Background

While many organizations in the private and public sectors are increasingly aware of the need to incorporate well-being measures as part of their efforts, the reality is that bottom line, quarterly-driven shareholder growth remains a dominant goal and metric. Short term growth is often the priority in the private sector and public sector. As long as organizations exist in a larger societal system which prioritizes financial success, these companies will remain under pressure to deliver financial results that do not fully incorporate societal and environmental impacts, measurements, or priorities.

Rather than focus solely on the negative aspects of how A/IS could harm humans and environments, we seek to explore how the implementation of well-being metrics can help A/IS to have a measurable, positive impact on human well-being as well as on systems and organizations. Incorporation of well-being goals and measures beyond what is strictly required can benefit both private sector organizations' brands and public sector organizations' stability and reputation, as well as help realize financial

Well-being

savings, innovation, trust, and many other benefits. For instance, a companion robot outfitted to support seniors in assisted living situations might traditionally be launched with a technology development model that was popularized by Silicon Valley known as "move fast and break things". The A/IS creator who rushed to bring the robot to market faster than the competition and who was unaware of well-being metrics, may have overlooked critical needs of the seniors. The robot might actually hurt the senior instead of helping by exacerbating isolation or feelings of loneliness and helplessness. While this is a hypothetical scenario, it is intended to demonstrate the value of linking A/IS design to well-being indicators.

By prioritizing largely fiscal metrics of success, A/IS devices might fail in the market because of limited adoption and subpar reception. However, if during use of the A/IS product, success were measured in terms of relevant aspects of well-being, developers and researchers could be in a better position to attain funding and public support. Depending on the intended use of the A/IS product, well-being measures that could be used extend to emotional levels of calm or stress; psychological states of thriving or depression; behavioral patterns of engagement in community or isolation; eating, exercise and consumption habits; and many other aspects of human well-being. The A/IS product could significantly improve quality of life guided by metrics from trusted sources, such as the World Health Organization, European Social Survey, and Sustainable Development Goal Indicators.

Thought leaders in the corporate arena have recognized the multifaceted need to utilize metrics beyond fiscal indicators.

PricewaterhouseCoopers defines "total impact" as a "holistic view of social, environmental, fiscal and economic dimensions—the big picture". Other thought-leading organizations in the public sector, such as the OECD, demonstrate the desire for business leaders to incorporate metrics of success beyond fiscal indicators for their efforts, exemplified in their 2017 workshop, Measuring Business Impacts on People's Well-Being. The B-Corporation movement has 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, B-Corps are defining their brands by provably aligning their efforts with wider measures of well-being.

Recommendations

A/IS creators should work to better understand and apply well-being metrics in the algorithmic age. Specifically:

- A/IS creators should work directly with experts, researchers, and practitioners in well-being concepts and metrics to identify existing metrics and combinations of indicators that would bring support a "triple bottom line", i.e., accounting for economic, social, and environmental impacts, approach to well-being. However, well-being metrics should only be used with consent, respect for privacy, and with strict standards for collection and use of these data.

- For A/IS to promote human well-being, the well-being metrics should be chosen in collaboration with the populations most affected by those systems—the A/IS

Well-being

stakeholders—including both the intended end-users or beneficiaries and those groups whose lives might be unintentionally transformed by them. This selection process should be iterative and through a learning and continually improving process. In addition, "metrics of well-being" should be treated as vehicles for learning and potential mid-course corrections. The effects of A/IS on human well-being should be monitored continuously throughout their life cycles, by A/IS creators and stakeholders, and both A/IS creators and stakeholders should be prepared to significantly modify, or even roll back, technology that is shown to reduce well-being, as defined by affected populations.

- A/IS creators in the business or academic, engineering, or policy arenas are advised to review the additional resources on standards development models and frameworks at the end of this chapter to familiarize themselves with existing indicators relevant to their work.

Further Resources

- PricewaterhouseCoopers (PwC). Managing and Measuring Total Impact: A New Language for Business Decisions, 2017.

- World Economic Forum. The Inclusive Growth and Development Report 2017, Geneva, Switzerland: World Economic Forum, January 16, 2017.

- OECD Guidelines on Measuring Subjective Well-being, 2013.

- National Research Council. Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience. DC: The National Academies Press, 2013.

Issue: A/IS creators have opportunities to safeguard human well-being by ensuring that A/IS does no harm to earth's natural systems or that A/IS contributes to realizing sustainable stewardship, preservation, and/or restoration of earth's natural systems. A/IS creators have opportunities to prevent A/IS from contributing to the degradation of earth's natural systems and hence losses to human well-being.

Background

It is unwise, and in truth impossible, to separate the well-being of the natural environment of the planet from the well-being of humanity. A range of studies, from the historic to more recent, prove that ecological collapse endangers human existence. Hence, the concept of well-being should encompass planetary well-being. Moreover, biodiversity and ecological integrity have intrinsic merit beyond simply their instrumental value to humans.

Technology has a long history of contributing to ecological degradation through its role in expanding the scale of resource extraction and environmental pollution, for example, the immense power needs of network computing, which leads to climate change, water scarcity, soil degradation, species extinction, deforestation,

Well-being

biodiversity loss, and destruction of ecosystems which in turn threatens humankind in the long run. These and other costs are often considered externalities and often do not figure into decisions or plans. At the same time, there are many examples, such as photovoltaics and smart grid technology that present potential ways to restore earth's ecosystems if undertaken within a systems approach aimed at sustainable economic and environmental development.

Environmental justice research demonstrates that the negative environmental impacts of technology are commonly concentrated on 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. Ecological impact can thus exacerbate the economic and sociological effects of wealth disparities on human well-being by concentrating environmental injustice onto those who are less well off. Moreover, well-being research findings indicate that unfair economic and social inequality has a dampening effect on everyone's well-being, regardless of economic or social class.

In these respects, A/IS are no exception; they can be used in ways that either help or harm the ecological integrity of the planet. It may be fair to say that ecological health and human well-being will, increasingly, depend upon A/IS creators. It is imperative that A/IS creators and stakeholders find ways to use A/IS to do no harm and to reduce the environmental degradation associated with economic growth—while simultaneously identifying applications to restore the ecological health of the planet and thereby safeguarding the well-being of humans. For A/IS to reduce environmental degradation and promote well-being, it is required that not only A/IS creators act along such lines, but also that a systems approach is taken by all A/IS stakeholders to find solutions that safeguard human well-being with the understanding that human well-being is inextricable from healthy social, economic, and environmental systems.

Recommendations

A/IS creators need to recognize and prioritize the stewardship of the Earth's natural systems to promote human and ecological well-being. Specifically:

- Human well-being should be defined to encompass ecological health, access to nature, safe climate and natural environments, biosystem diversity, and other aspects of a healthy, sustainable natural environment.

- A/IS systems should be designed to use, support, and strengthen existing ecological sustainability standards with a certification or similar system, e.g., LEED, Energy Star, or Forest Stewardship Council. This directs automation and machine intelligence to follow the principle of doing no harm and to safeguard environmental, social, and economic systems.

- A/IS creators should prioritize doing no harm to the Earth's natural systems, both intended and unintended harm.

- A committee should be convened to issue findings on ways in which A/IS can be used by business, NGOs, and governmental agencies to promote stewardship and restoration of natural systems while reducing the harmful impact of economic development on ecological sustainability and environmental justice.

Well-being

Further Resources

- D. Austin and M. Macauley. "Cutting Through Environmental Issues: Technology as a double-edged sword." The Brookings Institution, Dec. 2001 [Online]. Available: https://www.brookings.edu/articles/cutting-through-environmental-issues-technology-as-a-double-edged-sword/. [Accessed Dec. 1, 2018].

- J. Newton, *Well-being and the Natural Environment: An Overview of the Evidence.* August 20, 2007.

- P. Dasgupta, Human Well-Being and the Natural Environment. Oxford, U.K.: Oxford University Press, 2001.

- R. Haines-Young and M. Potschin. "The Links Between Biodiversity, Ecosystem Services and Human Well-Being," in *Ecosystem Ecology: A New Synthesis*, D. Raffaelli, and C. Frid, Eds. Cambridge, U.K.: Cambridge University Press, 2010.

- S. Hart, *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' of the Holy Father Francis On the Care for Our Common Home. May 24, 2015.

- "Environment," The 14th Dalai Lama. Accessed Dec. 9, 2018. https://www.dalailama.com/messages/environment.

- Why Islam.org, Environment and Islam, 2018.

Issue: Human rights law is related to, but distinct from, the pursuit of well-being. Incorporating a human-rights framework as an essential basis for A/IS creators means A/IS creators honor existing law as part of their well-being analysis and implementation.

Background

International human rights law has been firmly established for decades in order to protect various guarantees and freedoms as enshrined in charters such as the United Nations' Universal Declaration of Human Rights and the Council of Europe's Convention on Human Rights. In 2018, the Toronto Declaration on machine learning standards was released, calling on both governments and technology companies to ensure that algorithms respect basic principles of equality and non-discrimination. The Toronto Declaration sets forth an obligation to prevent machine learning systems from discriminating, and in some cases violating, existing human rights law.

Well-being initiatives are typically undertaken for the sake of public interest. However, any metric, including well-being metrics, can be misused to justify human rights violations. Encampment and mistreatment of refugees and ethnic cleansing undertaken to preserve a nation's culture (an aspect of well-being) is one example. Imprisonment or assassination of journalists or researchers to ensure the stability

Well-being

of a government is another. The use of well-being metrics to justify human rights violations is an unconscionable perversion of the nature of any well-being metric. It should be noted that these same practices happen today in relation to GDP. For instance, in 2012, according to the International Labour Organization (ILO), approximately 21 million people are victims of forced labor (slavery), representing 9% to 56% of GDP income for various countries. These clear human rights violations, from sex trafficking and use of children in armies, to indentured farming or manufacturing labor, can increase a country's GDP while obviously harming human well-being.

Well-being metrics are designed to measure the efficacy of efforts related to individual and societal flourishing. Well-being as a value complements justice, equality, and freedom. Well-designed application of well-being considerations by A/IS creators should not displace other issues of human rights or ethical methodologies, but rather complement them.

Recommendation

A human rights framework should represent the floor, and not the ceiling, for the standards to which A/IS creators must adhere. Developers and users of well-being metrics should be aware these metrics will not always adequately address human rights.

Further Resources

- United Nations Universal Declaration of Human Rights, 1948.

- Council of Europe's Convention on Human Rights, 2018.

- International Labor Organization (ILO) Declaration on Fundamental Principles and Rights at Work, 1998.

- The regularly updated University of Minnesota Human Rights Library provides a wealth of material on human rights laws, its history, and the organizations engaged in promoting them.

- The Oxford Human Rights Hub reports on how and why technologies surrounding artificial intelligence raise human rights issues.

Well-being

Section 2—Implementing Well-being Metrics for A/IS Creators

A key challenge for A/IS creators in realizing the benefits of well-being metrics is how to best incorporate them into their work. This section explores current best thinking on how to make this happen.

Issue: How can A/IS creators incorporate well-being into their work?

Background

Without practical ways of incorporating well-being metrics to guide, measure, and monitor impact, A/IS will likely lack fall short of its potential to avoid harm and promote well-being. Incorporating well-being thinking into typical organizational processes of design, prototyping, marketing, etc., suggests a variety of adaptations.

Organizations and A/IS creators should consider clearly defining the type of A/IS product or service that they are developing, including articulating its intended stakeholders and uses. By defining typical uses, possible uses, and finally unacceptable uses of the technology, creators will help to spell out the context of well-being. This can help to identify possible harms and risks given the different possible uses and end users, as well as intended and unintended positive consequences.

Additionally, internal and external stakeholders should be extensively consulted to ensure that impacts are thoroughly considered through an iterative and learning stakeholder engagement process. After consultation, A/IS creators should select appropriate well-being indicators based on the possible scope and impact of their A/IS product or service. These well-being indicators can be drawn from mainstream sources and models and adapted as necessary. They can be used to engage in pre-assessment of the intended user population, projection of possible impacts, and post-assessment. Development of a well-being indicator measurement plan and relevant data infrastructure will support a robust integration of well-being. A/IS models can also be trained to explicitly include well-being indicators as subgoals.

Data and discussions on well-being impacts can be used to suggest improvements and modifications to existing A/IS products and services throughout their lifecycle. 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, increase their sense of isolation, and lead to an overall decrease in their well-being. Therefore, even though a product modification may increase well-being according to one indicator or set of

Well-being

A/IS stakeholders, it does not mean that this modification should automatically be adopted.

Finally, organizational processes can be modified to incorporate the above strategies. Appointment of an organizational lead person for well-being impacts, e.g., a well-being lead, ombudsman, or officer can help to facilitate this effort.

Recommendation

A/IS creators should adjust their existing development, marketing, and assessment cycles to incorporate well-being concerns throughout their processes. This includes identification of an A/IS lead ombudsperson or officer; identification of stakeholders and end users; determination of possible uses, harm and risk assessment; robust stakeholder engagement; selection of well-being indicators; development of a well-being indicator measurement plan; and ongoing improvement of A/IS products and services throughout the lifecycle.

Further Resources

- Peter Senge and the Learning Organization - (synopsis) Purdue University

- Stakeholder Engagement: A Good Practice Handbook for Companies Doing Business in Emerging Markets. International Finance Corporation, May 2007.

- Global Reporting Initiative

- GNH Certification, Centre for Bhutan and GNH Studies, 2018.

- J. Helliwell, R. Layard, and J. Sachs, Eds., "The Objective Benefits of Subjective Well-Being," in World Happiness Report 2013. New York: UN Sustainable Development Solutions Network, pp. 54-79, 2013.

- Global Happiness and Well-being Policy Report by the Global Happiness Council, 2018.

Issue: How can A/IS creators influence A/IS goals to ensure well-being, and what can A/IS creators learn or borrow from existing models in the well-being and other arenas?

Background

Another way to incorporate considerations of well-being is to include well-being measures in the development, goal setting, and training of the A/IS systems themselves.

Identified metrics of well-being could be formulated as auxiliary objectives of the A/IS. As these auxiliary well-being objectives will be only a subset of the intended goals of the system, the architecture will need to balance multiple objectives. Each of these auxiliary objectives may be expressed as a goal, set of rules, set of values, or as a set of preferences, which can be weighted and combined using established methodologies from intelligent systems engineering.

Well-being

For example, an educational A/IS tool could not only optimize learning outcomes, but also incorporate measures of student social and emotional education, learning, and thriving.

A/IS-related data relates both to the individual—through personalized algorithms, in conjunction with affective sensors measuring and influencing emotion, and other aspects of individual well-being—and to society as large data sets representing aggregate individual subjective and objective data. As the exchange of this data becomes more widely available via establishing tracking methodologies, the data can be aligned within A/IS products and services to increase human well-being. For example, 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 Pepper units (and ostensibly their owners) benefit as well. As long as this data exchange happens with the predetermined consent of the robots' owners, this innovation in real time model can be emulated for the large-scale aggregation of information relating to existing well-being metrics.

A/IS creators can also help to operationalize well-being metrics by providing stakeholders with reports on the expected or actual outcomes of the A/IS and the values and objectives embedded in the systems. This transparency will help creators, users, and third parties assess the state of well-being produced by A/IS and make improvements in A/IS. In addition, A/IS creators should consider allowing end users to layer on their own preferences, such as allowing users

to limit their use of an A/IS product if it leads to increased sustained stress levels, sustained isolation, development of unhealthy habits, or other decreases to well-being.

Incorporating well-being goals and metrics into broader organizational values and processes would support the use of well-being metrics as there would be institutional support. A key factor in industrial, corporate, and societal progress is cross-dissemination of concepts and models from one industry or field to another. To date, a number of successful concepts and models exist in the fields of sustainability, economics, industrial design and manufacturing, architecture and urban development, and governmental policy. These concepts and models can provide a foundation for building a metrics standard and the use of well-being metrics by A/IS creators, from conception and design to marketing, product updates, and improvements to the user experience.

Recommendation

Create technical standards for representing goals, metrics, and evaluation guidelines for well-being metrics and their precursors and components within A/IS that include:

- Ontologies for representing technological requirements.

- A testing framework for validating adherence to well-being metrics and ethical principles such as IEEE P7010™ Standards Project for Well-being Metric for Autonomous and Intelligent Systems.

Well-being

- The exploration of models and concepts listed above as well as others as a basis for a well-being metrics standard for A/IS creators. *(See page 191, Additional Resources: Additional Resources: Standards Development Models and Frameworks)*

- The development of a well-being metrics standard for A/IS that encompasses an understanding of well-being as holistic and interlinked to social, economic, and ecological systems.

Further Resources

- A.F.T 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

- R. A. Calvo, and D. Peters. Positive Computing: Technology for Well-Being and Human Potential. Cambridge MA: MIT Press, 2014.

- Y. Collette, and P. Slarry. Multiobjective Optimization: Principles and Case Studies (Decision Engineering Series). Berlin, Germany: Springer, 2004. doi: 10.1007/978-3-662-08883-8.

- J. Greene, 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.

- L. Li, I. Yevseyeva, V. Basto-Fernandes, H. Trautmann, N. Jing, and M. Emmerich,"Building and Using an Ontology of Preference-Based Multiobjective Evolutionary Algorithms." In 9th International Conference on Evolutionary Multi-Criterion Optimization—Volume 10173 (EMO

2017), H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecek, Y. Jin, and C. Grimme, Eds., Vol. 10173. Springer-Verlag, Berlin, Heidelberg, 406-421, 2017.

- PositiveSocialImpact: Empowering people, organizations and planet with information and knowledge to make a positive impact to sustainable development, 2017.

- D.K. Ura, Bhutan's Gross National Happiness Policy Screening Tool.

Issue: Decision processes for determining relevant well-being indicators through stakeholder deliberations need to be established.

Background

A/IS stakeholder involvement is necessary to determine relevant well-being indicators, for a number of reasons:

- "Well-being" will be defined differently by different groups affected by A/IS. The most relevant indicators of well-being may vary according to country, with concerns of wealthy nations being different than those of low- and middle-income countries. Indicators may vary based on geographical region or unique circumstances. The indicators may also be different across social groups, including gender, race, ethnicity, and disability status.

- Common indicators of well-being include satisfaction with life, healthy life expectancy,

Well-being

economic standard of living, trust in government, social support, perceived freedom to make life decisions, income equality, access to education, and poverty rates. Applying them in particular settings necessarily requires judgment, to ensure that assessments of well-being are in fact meaningful in context and reflective of the life circumstances of the diverse groups in question.

- Not all aspects of well-being are easily quantifiable. The importance of hard-to-quantify aspects of well-being is most likely to become apparent through interaction with those more directly affected by A/IS in specific settings.

- Engineers and corporate employees frequently misunderstand stakeholders' needs and expectations, especially when the stakeholders are very different from them in terms of educational and cultural background, social location, and/or economic status.

The processes through which stakeholders become involved in determining relevant well-being indicators will affect the quality of the indicators selected and assessed. Stakeholders should be empowered to define well-being, assess the appropriateness of existing indicators and propose new ones, and highlight context-specific factors that bear on issues of well-being, whether or not the issues have been recognized previously or are amenable to measurement. Interactive, open-ended discussions or deliberations among a wide variety of stakeholders and system designers are more likely to yield robust, widely-shared understandings of well-being and how to measure it in context. Closed-ended or over-determined methods for soliciting stakeholder input are likely to miss relevant information that system designers have not anticipated.

A process of stakeholder engagement and deliberation is one model for collective decision-making. Parties in such deliberation come together as equals. Their goal is to set aside their immediate, personal interests in order to think together about the common good. Participants in a stakeholder engagement and deliberation learn from one another's perspectives and experiences.

In the real world, stakeholder engagement and deliberation may run into the following challenges:

- Individuals with more education, power, or higher social status may—intentionally or unintentionally—dominate the discussion, undermining their ability to learn from less powerful participants.

- Topics may be preemptively ruled "out of bounds", to the detriment of collective problem-solving. An example would be if, in a deliberation on well-being and A/IS, participants were told that worries about the costs of health insurance were unrelated to A/IS and thus could not be discussed.

- Engineers and scientists may claim authority over technical issues and be willing to deliberate only on social issues, obscuring the ways that technical and social issues are intertwined.

- Less powerful groups may be unable to keep more powerful ones "at the table" when discussions get contentious, and vice versa.

- Participants may not agree on who can legitimately be involved in the conversation. For example, the consensual spirit of deliberation is often used as a justification for excluding activists and others who already hold a position on the issue.

Well-being

Stakeholder engagement and deliberative processes can be effective when:

- Their design is guided by experts or practitioners who are experienced in deliberation models.

- Deliberations are facilitated by individuals sensitive to issues of power and are skilled in mediating deliberation sessions.

- Less powerful actors participate with the help of allies who can amplify their voices.

- More powerful actors participate with an awareness of their own power and make a commitment to listen with humility, curiosity, and open-mindedness.

- Deliberations are convened by institutions or individuals who are trusted and respected by all parties and who hold all actors accountable for participating constructively.

Ethically aligned design of A/IS would be furthered by thoughtfully constructed, context-specific deliberations on well-being and the best indicators for assessing it.

Recommendation

Appoint a lead team or person, "leads", to facilitate stakeholder engagement and to serve as a resource for A/IS creators who use stakeholder-based processes to establish well-being indicators. Specifically:

- Leads should solicit and collect lessons learned from specific applications of stakeholder engagement and deliberation in order to continually refine its guidance.

- When determining well-being indicators, the leads should enlist the help of experts in public participation and deliberation. With expert guidance, facilitators can provide guidance for how to: take steps to mitigate the effects of unequal power in deliberative processes; incorporate appropriately trained facilitators and coaching participants in deliberations; recognize and curb disproportionate influence by more-powerful groups; use techniques to maximize the voices of less-powerful groups.

- Leads should use their convening power to bring together A/IS creators and stakeholders, including critics of A/IS, for deliberations on well-being indicators, impacts, and other considerations for specific contexts and settings. Leads' involvement would help bring actors to the table with a balance of power and encourage all actors to remain in conversation until robust, mutually agreeable definitions are found.

Further Resources

- D. E. Booher and J. E. Innes. Planning with Complexity: An Introduction to Collaborative Rationality for Public Policy. London: Routledge, 2010.

- J. A. Leydens and J. C. Lucena. Engineering Justice: Transforming Engineering Education and Practice. Wiley-IEEE Press, 2018.

- G. Ottinger. Assessing Community Advisory Panels: A Case Study from Louisiana's Industrial Corridor. Center for Contemporary History and Policy, 2008.

- Expert and Citizen Assessment of Science and Technology (ECAST) Network

Well-being

Issue: There are insufficient mechanisms to foresee and measure negative impacts, and to promote and safeguard positive impacts of A/IS.

Background

A/IS technologies present great opportunity for positive change in every aspect of society. However, they can—by design or unintentionally—cause harm as well. While it is important to consider and make sense of possible benefits, harms, and trade-offs, it is extremely challenging to foresee all of the relevant, direct, and secondary impacts.

However, it is prudent to review case studies of similar products and the impacts they have had on well-being, as well as to consider possible types of impacts that could apply. Issues to consider include:

- Economic and labor impacts, including labor displacement, unemployment, and inequality,
- Accountability, transparency, and explainability,
- Surveillance, privacy, and civil liberties,
- Fairness, ethics, and human rights,
- Political manipulation, deception, "nudging", and propaganda,
- Human physical and psychological health,
- Environmental impacts,
- Human dignity, autonomy, and human vs. A/IS roles,
- Security, cybersecurity, and autonomous weapons, and
- Existential risk and super intelligence.

While this is a partial list, it is important to be aware of and reflect on possible and actual cases. For example:

- A prominent concern related to A/IS is of labor displacement and economic and social impacts at an individual and a systems level. A/IS technologies designed to replicate human tasks, behavior, or emotion have the potential to increase or decrease human well-being. These systems could complement human work and increase productivity, wages, and leisure time; or they could be used to supplement and displace human workers, leading to unemployment, inequality, and social strife. It is important for A/IS creators to think about possible uses of their technology and whether they want to encourage or design in restrictions in light of these impacts.

- Another example relates to manipulation. Sophisticated manipulative technologies utilizing A/IS can restrict the fundamental freedom of human choice by manipulating humans who consume content without them recognizing the extent of the manipulation. Software platforms are moving from targeting and customizing content to much more powerful and potentially harmful "persuasive computing" that leverages psychological data and methods. While these approaches may be effective in encouraging use of a product, they may come at significant psychological and social costs.

- A/IS may 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 intelligence from artificial intelligence, there is a significant risk

Well-being

that unscrupulous operators will abuse the technology for unethical commercial or outright criminal purposes. 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 and via multiple mediums.

A potential entry point for exploring these unintended consequences is computational sustainability.

Computational-Sustainability.org defines the term as an "interdisciplinary field that aims to apply techniques from computer science, information science, operations research, applied mathematics, and statistics for balancing environmental, economic, and societal needs for sustainable development". The Institute of Computational Sustainability states that the intent of computational sustainability is provide "computational models for a sustainable environment, economy, and society". Examples of applied computational sustainability can be seen in the Stanford University Engineering Department's course in computational sustainability presentation. Computational sustainability technologies designed to increase social good could also be tied to existing well-being metrics.

Recommendation

- To avoid potential negative, unintended consequences, and secure and safeguard positive impacts, A/IS creators, end-users, and stakeholders should be aware of possible well-being impacts when designing, using, and monitoring A/IS systems. This includes being aware of existing cases and possible areas of impact, measuring impacts on well-being outcomes, and developing regulations to promote beneficent uses of A/IS. Specifically:

- A/IS creators should protect human dignity, autonomy, rights, and well-being of those directly and indirectly affected by the technology. As part of this effort, it is important to include multiple stakeholders, minorities, marginalized groups, and those often without power or a voice in consultation.

- Policymakers, regulators, monitors, and researchers should consider issuing guidance on areas such as A/IS labor and the proper role of humans vs. A/IS in work transparency, trust, and explainability; manipulation and deception; and other areas that emerge.

- Ongoing literature review and analysis should be performed by research and other communities to curate and aggregate information on positive and negative A/IS impacts, along with demonstrated approaches to realize positive ones and ameliorate negative ones.

- A/IS creators working toward computational sustainability should integrate well-being concepts, scientific findings, and indicators into current computational sustainability models. They should work with well-being experts, researchers, and practitioners to conduct research and develop and apply models in A/IS development that prioritize and increase human well-being.

Well-being

- Cross-pollination should be developed between computational sustainability and well-being professionals to ensure integration of well-being into computational sustainability frameworks, and vice versa. Where feasible and reasonable, do the same for conceptual models such as doughnut economics and systems thinking.

Further Resources

- AI Safety Research by The Future of Life Institute

- D. Helbing, et al. "Will Democracy Survive Big Data and Artificial Intelligence?" *Scientific American*, February 25, 2017.

- J. L. Schenker, "Can We Balance Human Ethics with Artificial Intelligence?" *Techonomy,* January 23, 2017.

- M. Bulman, "EU to Vote on Declaring Robots To Be 'Electronic Persons'." *Independent,* January 14, 2017.

- N. Nevejan, for the European Parliament. "European Civil Law Rules in Robotics." October 2016.

- University of Oxford. "Social media manipulation rising globally, new report warns," https://phys.org/news/2018-07-social-media-globally.html. July 20, 2018.

- "The AI That Pretends To Be Human," *LessWrong* blog post, February 2, 2016.

- C. Chan, "Monkeys Grieve When Their Robot Friend Dies." *Gizmodo*, January 11, 2017.

- Partnership on AI, "AI, Labor, and the Economy" Working Group launches in New York City," https://www.partnershiponai.org/aile-wg-launch/. April 25, 2018.

- C.Y. Johnson, "Children can be swayed by robot peer pressure,study says," The Washington Post, August 15, 2018. [Online]. Available: www.WashingtonPost.com. [Accessed 2018].

Further Resources for Computational Sustainability

- Stanford Engineering Department, Topics in Computational Sustainability Course Presentation, 2016.

- Computational Sustainability, Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society Project Summary.

- C. P. Gomes, "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.

- S.J. Gershman, E. J. Horvitz, and J. B. Tenenbaum. "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines," *Science* vol. 349, no. 6245, pp. 273–278, July 2015.

- ACM Fairness, Accountability and Transparency Conference

Well-being

Thanks to the Contributors

We wish to acknowledge all of the people who contributed to this chapter.

The Well-being Committee

- **John C. Havens** (Co-Chair) – Executive Director, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems; Executive Director, The Council on Extended Intelligence; Author, *Heartificial Intelligence: Embracing Our Humanity to Maximize Machines*

- **Laura Musikanski** (Co-Chair) – Executive Director at The Happiness Alliance—home of The Happiness Initiative & Gross National Happiness Index

- **Liz Alexander** – PhD Futurist

- **Anna Alexandrova** – Senior Lecturer in Philosophy of Science at Cambridge University and Fellow of Kings College

- **Christina Berkley** – Executive Coach to leaders in exponential technologies, cutting-edge science, and aerospace

- **Catalina Butnaru** – UK AI Ambassador for global community City.AI, and Founder of HAI, the first methodology for applications of AI in cognitive businesses

- **Celina Beatriz** – Project Director at the Institute for Technology & Society of Rio de Janeiro (ITS Rio)

- **Peet van Biljon** – Founder and CEO at BMNP Strategies LLC, advisor on strategy, innovation, and business transformation; Adjunct faculty at Georgetown University; Business ethics author

- **Amy Blankson** – Author of The Future of Happiness and Founder of TechWell, a research and consulting firm that aims to help organizations to create more positive digital cultures

- **Marc Böhlen** – Professor, University at Buffalo, Emerging Practices in Computational Media. www.realtechsupport.org

- **Rafael A. Calvo** – Professor and ARC Future Fellow at The University of Sydney. Co-author of Positive Computing: Technology for Well-Being and Human Potential

- **Rumman Chowdhury** – PhD Senior Principal, Artificial Intelligence, and Strategic Growth Initiative Responsible AI Lead, Accenture

- **Dr. Aymee Coget** – CEO and Founder of Happiness For HumanKind

- **Danny W. Devriendt** – Managing director of Mediabrands Dynamic (IPG) in Brussels, and the CEO of the Eye of Horus, a global think-tank for communication-technology related topics

- **Eimear Farrell** – Eimear Farrell, independent expert/consultant on technology and human rights (formerly at OHCHR)

- **Danit Gal** – Project Assistant Professor, Keio University; Chair, IEEE Standard P7009 on the Fail-Safe Design of Autonomous and Semi-Autonomous Systems

Well-being

- **Marek Havrda** – PhD Strategy Advisor, GoodAI
- **Andra Keay** – Managing Director of Silicon Valley Robotics, cofounder of Robohub
- **Dr. Peggy Kern** – Senior Lecturer, Centre for Positive Psychology at the University of Melbourne's Graduate School of Education
- **Michael Lennon** – Senior Fellow, Center for Excellence in Public Leadership, George Washington University; Co-Founder, Govpreneur.org; Principal, CAIPP.org (Consortium for Action Intelligence and Positive Performance); Member, Well-being Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems Committee
- **Alan Mackworth** – Professor of Computer Science, University of British Columbia; Former President, AAAI; Co-author of "Artificial Intelligence: Foundations of Computational Agents"
- **Richard Mallah** – Director of AI Project, Future of Life Institute
- **Fabrice Murtin** – Senior Economist, OECD Statistics and Data Directorate
- **Gwen Ottinger** – Associate Professor, Center for Science, Technology, and Society and Department of Politics, Drexel University; Director, Fair Tech Collective
- **Eleonore Pauwels** – Research Fellow on AI and Emerging Cybertechnologies, United Nations University (NY) and Director of the AI Lab, Woodrow Wilson International Center for Scholars (DC)
- **Venerable Tenzin Priyadarshi** – MIT Media Lab, Director, Ethics Initiative

- **Gideon Rosenblatt** – Writer, focused on work and the human experience in an era of machine intelligence, at The Vital Edge
- **Daniel Schiff** – PhD Student, Georgia Institute of Technology; Chair, Sub-Group for Autonomous and Intelligent Systems Implementation, IEEE P7010™ Standards Project for Well-being Metric for Autonomous and Intelligent Systems
- **Madalena Sula** – Undergraduate student of Electrical and Computer Engineering, University of Thessaly, Greece, x-PR Manager of IEEE Student Branch of University of Thessaly, Data Scientist & Business Analyst in a multinational company
- **Vincent Siegerink** – Analyst, OECD Statistics and Data Directorate
- **Andy Townsend** – Emerging and Disruptive Technology, PwC UK
- **Andre Uhl** – Research Associate, Director's Office, MIT Media Lab
- **Ramón Villasante** – Founder of PositiveSocialImpact. Software designer, engineer, CTO & CPO in EdTech for sustainable development, social impact and innovation
- **Sarah Villeneuve** – Policy Analyst; Member, IEEE P7010™ Standards Project for Well-being Metric for Autonomous and Intelligent Systems.

For a full listing of all IEEE Global Initiative Members, visit standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_bios.pdf.

For information on disclaimers associated with EAD1e, see *How the Document Was Prepared*.