

Affective Computing

Affect is a core aspect of intelligence. Drives and emotions, such as excitement and depression, are used to coordinate action throughout intelligent life, even in species that lack a nervous system. We are coming to realize that emotions are not an impediment to rationality, arguably they are integral to rationality in humans. Emotions are one evolved mechanism for satisficing – for getting what needs to be done in the time available with the information at hand. Emotions are core to how individuals and societies coordinate their actions. Humans are therefore susceptible to emotional influence both positively and negatively.

We would like to ensure that AI will be used to help humanity to the greatest extent possible in all contexts. In particular, artifacts used in society could cause harm either by amplifying or damping human emotional experience. It is quite possible we have reached a point where AI is affecting humans psychologically more than we realize. Further, even the rudimentary versions of synthetic emotions already in use have significant impact on how AI is perceived by policy makers and the general public.

This subcommittee addresses issues related to emotions and emotion-like control in both humans and artifacts. Our working groups have put forward candidate recommendations on a variety of concerns: considering how affect varies across human cultures, the particular problems of artifacts designed for intimate relations, considerations of how intelligent artifacts may be used for “nudging,” how systems can support (or at least not interfere with) human flourishing, and appropriate policy concerning artifacts designed with their own affective systems.

Affective Computing

Document Sections

- [Systems Across Cultures](#)
- [When Systems Become Intimate](#)
- [System Manipulation/Nudging/Deception](#)
- [Systems Supporting Human Potential \(Flourishing\)](#)
- [Systems with Their Own Emotions](#)

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.

Affective Computing

Systems Across Cultures

Issue:

Should affective systems interact using the norms appropriate for verbal and nonverbal communication consistent with the societal norms where they are located?

Background

Societies and therefore individuals around the world have different ways to maintain eye contact, express intentions through gestures, interpret silence, etc. These particularities could be incorporated into the affective systems in order to transmit the intended message. It would seem that an extensive study surrounding the norms/values of the community where the affective system will be deployed is essential to the system acceptability.

Candidate Recommendations

Any successful affective system should have a minimum set of ethical values/norms in its knowledge base that should be used in a specific cultural context. Some examples are listed below:

1. Affective systems should be careful in using small talk. Although small talk is useful for acting friendly, some communities see people that use small talk as insincere and hypocritical, while other cultures see the opposite and tend to consider people that do not use small talk as unfriendly, uncooperative, rude, arrogant, or ignorant. Additionally speaking with proper vocabulary, grammar, and sentence structure is often in contrast to the typical interactions that people have. In many mature economies, the latest trend, TV show, or other media can significantly influence what is viewed as appropriate vocabulary and interaction style.
2. Affective systems should recognize that the amount of personal space (proxemics) given is very important for human interaction. People from different cultures have different comfort zone distances to establish smooth communication. Crossing these limits without permission can transmit negative messages, such as hostile or sexual overtures.
3. Eye contact is an essential component in social interaction for certain cultures, while for others, it is not essential and may even generate misunderstandings or conflicts. It is important to recognize this in the development of such systems.

Affective Computing

4. Hand gestures and other non-verbal interaction are very important for social interaction, but should be used with caution across cultures and should be acknowledged in the design of affective systems. For instance, although a “thumbs-up” sign is commonly used to indicate approval, in some countries this gesture can be considered an insult.
 5. Facial expressions are often used to detect emotions and facilitate emotional conversations. While it is tempting to develop A/IS that can recognize, analyze, and even display facial expressions for social interaction, it should be noted that facial expressions may not be universal across cultures and that an AI system trained with a dataset from one culture may not be readily usable in another culture.
- Elmer, D. *Cross-Cultural Connections: Stepping Out and Fitting In Around the World*. Westmont, IL: InterVarsity Press, 2002.
 - Price, M. “Facial Expressions—Including Fear—May Not Be as Universal as We Thought.” *Science*, October 17, 2016.

Issue:
Long-term interaction with affective artifacts lacking cultural sensitivity could alter the way people interact in society.

Further Resources

The following documents/organizations can be used as additional resources to support the development of ethical affective systems.

- Cotton, G. “[Gestures to Avoid in Cross-Cultural Business: In Other Words, ‘Keep Your Fingers to Yourself!’](#)” *Huffington Post*, June 13, 2013.
- “[Cultural Intelligence & Paralanguage: Using Your Voice Differently Across Cultures.](#)” Sydney, Aus.: Culture Plus Consulting, 2016.
- Cotton, G. *Say Anything to Anyone, Anywhere: 5 Keys To Successful Cross-Cultural Communication*. Hoboken, NJ: Wiley, 2013.

Background

Systems that do not have cultural knowledge incorporated into their knowledge base may change the way people interact, which may impact not only individuals, but also an entire society. Humans often use mirroring in order to understand and develop their principles and norms for behavior. At the same time, certain machine learning approaches focus on how to more appropriately interact with humans by mirroring human behavior. So learning via mirroring can go in both directions. If affective artifacts without cultural sensitivity interact with impressionable humans, they could alter the norms, principles, and therefore actions of that person. This creates a situation where the impact

Affective Computing

of interacting with machines could significantly alter societal and cultural norms. For instance, children interacting with these systems can learn social and cultural values, which may be different from those present in their local community.

Candidate Recommendations

1. It is necessary to survey and analyze the long-term interaction of people with affective systems with different protocols and metrics to measure the modifications of habits, norms, and principles as well as the cultural and societal impacts.
2. Responsible parties (e.g., parents, nurse practitioners, social workers, and governments) should be trained to detect the influence due to AI and in effective mitigation techniques. In the most extreme case it should always be possible to shut down harmful A/IS.

Further Resources

The following documents can be used as guides to support the development of ethical affective systems.

- Nishida, T., and C. Faucher. *Modelling Machine Emotions for Realizing Intelligence: Foundations and Applications*. Berlin, Germany: Springer-Verlag, 2010.
- Pauleen, D. J. et al. "Cultural Bias in Information Systems Research and Practice: Are You Coming From the Same Place I Am?" *Communications of the Association for Information Systems* 17 (2006): 1–36.
- Bielby, J. "Comparative Philosophies in Intercultural Information Ethics." *Confluence: Online Journal of World Philosophies* 2, no. 1 (2015): 233–253.
- Bryson, J., "[Why Robot Nannies Probably Won't Do Much Psychological Damage.](#)" A commentary on an article by N. Sharkey and A. Sharkey, *The Crying Shame of Robot Child Care Companions*. *Interaction Studies* 11, no. 2 (2010): 161–190.
- Sharkey, A., and N. Sharkey. "Children, the Elderly, and Interactive Robots." *IEEE Robotics & Automation Magazine* 18, no. 1 (2011): 32–38.

Issue:

When affective systems are inserted across cultures, they could affect negatively the cultural/socio/religious values of the community where they are inserted.

Background

Some philosophers believe there are no universal ethical principles; instead they argue that ethical norms vary from society to society. Regardless of whether universalism or some form of ethical relativism is true, affective systems need to respect the values of the cultures within which

Affective Computing

they are embedded. To some it may be that we should be designing affective systems which can reflect the values of those with which the systems are interacting. There is a high likelihood that when spanning different groups, the values imbued by the developer will be different from the operator or customer of that affective system. Differences between affective systems and societal values can generate conflict situations (e.g., gestures being misunderstood, or prolonged or inadequate eye contact) that may produce undesirable results, perhaps even physical violence. Thus, affective systems should adapt to reflect the values of the community (and individuals) where they will operate in order to avoid conflict.

Candidate Recommendation

Assuming the affective systems have a minimum subset of configurable ethical values incorporated in their knowledge base:

1. They should have capabilities to identify differences between their values and the values of those they are interacting with and alter their interactions accordingly. As societal values change over time, any affective system needs to have the capability to detect this evolution and adapt its current ethical values to be in accordance with other people's values.
2. Those actions undertaken by an affective system that are most likely to generate

an emotional response should be designed to be easily changed. Similar to how software today externalizes the language and vocabulary to be easily changed based on location, affective systems should externalize some of the core aspects of their actions.

Further Resources

The following documents/organizations can be used as guides to support the development of ethical affective systems.

- Bielby, J. "Comparative Philosophies in Intercultural Information Ethics." *Confluence: Online Journal of World Philosophies* 2, no. 1 (2015): 233–253.
- Velasquez, M., C. Andre, T. Shanks, and M. J. Meyer. "[Ethical Relativism](#)." Markkula Center for Applied Ethics, Santa Clara, CA: Santa Clara University, August 1, 1992.
- Culture reflects the moral values and ethical norms governing how people should behave and interact with others. "[Ethics, an Overview](#)." Boundless Management.
- Donaldson, T. "[Values in Tension: Ethics Away from Home](#)." *Harvard Business Review*. September–October 1996.
- [The Center for Nonviolent Communication](#).

Affective Computing

When Systems Become Intimate

Issue:

Are moral and ethical boundaries crossed when the design of affective systems allows them to develop intimate relationships with their users?

Background

While robots capable of participating in an intimate relationship are not currently available, the idea that they could become intimate sexual partners with humans (e.g., sex robots) is one that captures the attention of the public and the media. Because the technology is already drawing much ethical scrutiny and may raise significant ethical concerns, it is important that policy makers and the professional community participate in developing guidelines for ethical research in this area. Part of the goal is to highlight potential ethical benefits and risks that may emerge, if and when affective systems develop intimacy with their users. Robots for use in the sex industry may help lessen human trafficking and the spread of STIs, but there is also the possibility that these systems could negatively impact human-to-human intimate relations. Human-to-human relations are currently viewed as being more rewarding, but also much more difficult to maintain than, for example, use of future robotic sex workers.

Candidate Recommendation

As this technology develops, it is important to monitor research in this realm and support those projects that enhance the user's development of intimate relationships in positive and therapeutic ways while critiquing those that contribute to problematic intimate relations, specifically:

1. Intimate systems must not be designed or deployed in ways that contribute to sexism, negative body image stereotypes, gender or racial inequality.
2. Intimate systems must avoid the sexual/psychological manipulation of the users of these systems unless the user is made aware they are being manipulated in this way (opt-in).
3. Intimate systems should not be designed in a way that contributes to user isolation from other human companions.
4. Designers of affective robotics, especially intimate systems, must foresee and publicly acknowledge that these systems can interfere with the relationship dynamics between human partners, causing jealousy or feelings of disgust to emerge between human partners.
5. Intimate systems must not foster deviant or criminal behavior. Sex robots should not be built in ways that lead to the normalization of taboo, unethical, or illegal sexual practices, such as pedophilia or rape.

Affective Computing

6. Commercially marketed AI should not be considered to be a person in a legal sense, nor marketed as a person. Rather its artificial (authored, designed, and built deliberately) nature should always be made as transparent as possible, at least at point of sale and in available documentation (as noted in the Systems Supporting Human Potential Section below).

Further Resources

The following documents/organizations are provided for further research.

- Levy, D. [Love and Sex with Robots: The Evolution of Human-Robot Relationships](#). New York: HarperCollins Publishers, 2007
- Scheutz, M. "The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots," in *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. Bekey, 205. Cambridge, MA: MIT Press, 2011.
- Richardson, K. "The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots." *ACM SIGCAS Newsletter, SIGCAS Computers & Society* 45, no. 3 (2015): 290–293.
- Sullins, J. P. "Robots, Love, and Sex: The Ethics of Building a Love Machine." *IEEE Transactions on Affective Computing* 3, no. 4 (2012): 398–409.
- Yeoman, I., and M. Mars. "[Article Robots, Men and Sex Tourism](#)." *Futures* 44, no. 4 (2012): 365–371.
- [Campaign against Sex Robots](#).
- Whitby, B. "Do You Want a Robot Lover? The Ethics of Caring Technologies," in *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. A. Bekey, 233–248. Cambridge, MA: MIT Press, 2012.
- Danaher, J., and N. McArthur. *Robot Sex: Sexual and Ethical Implications*. Cambridge, MA: MIT Press, 2017.

Issue:

Can and should a ban or strict regulations be placed on the development of sex robots for private use or in the sex industry?

Background

The very idea of sex robots has sparked controversy even before many of these systems have become available. At this time, sex robots tend to be expensive love dolls made of silicone placed over a metal skeleton. These dolls can include robotic systems such as heating elements, sensors, movement, and rudimentary AI. The current state of the technology is a far cry from the sex robots imagined in novels and other media but they may just be the first step toward more advanced systems. There is ongoing debate around these systems. Critics are calling for strict

Affective Computing

regulation or even a full ban on the development of this technology, while others argue that social value could be found by developing intimate robots, including on religious grounds.

Sex robots are already used for prostitution and this trend is likely to continue in many regions of the world. Some researchers report that robot prostitutes will completely revolutionize the sex tourism industry by 2050. For example, by that time, Amsterdam's Red Light District may be dominated by a variety of android systems with various capabilities (Yeoman and Mars, 2012). However there are critics of the technology, including those who are calling for an outright ban.

Despite being illegal, prostitution commonly occurs in many societies. Yet it is rarely done without creating deep ethical problems for the sex workers themselves and the societies in which they operate. Sex robots may alleviate some of these ethical concerns; for instance it has been argued that:

1. Sex robots might be less likely to be a vector for the transmission of sexually transmitted infections (STIs).
2. Sex robots could greatly lessen human trafficking of sex workers.
3. Sex robots could be regulated by policies on controlling prices, hours of operations, sexual services, and other aspects of prostitution.

However the technology can create serious ethical problems such as:

1. This technology would likely further normalize the sex industry, and that typically

means a further tendency to objectify women, given that the majority of clients for these technologies are heterosexual men.

2. The availability of the technology could disrupt intimate relationships between human beings.

Human sexuality is an important human activity, but it comes associated with difficult ethical issues related to power and desire. This means that robot sexual partners will always be an ethically contentious technology. A comprehensive/global ban on sex robots is unlikely given that a large market for these technologies may already exist and is part of the current demand for sex toys and devices. However, there are important issues/considerations that the designers of these technologies need to consider.

Candidate Recommendation

1. We recommend regulation, not a ban, in accordance with cultural norms.
2. Existing laws regarding personal imagery need to be reconsidered in light of robot sexuality.
3. If it is proven through scientific studies that therapeutic uses of this technology could reduce recidivism in those who commit sex crimes, controlled use for those purposes only should be permitted, under legal and/or medical supervision.
4. Robot prostitution and sex tourism need to be monitored and controlled to fit local laws and policies.

Affective Computing

Further Resources

- Danaher, J., and N. McArthur. *Robot Sex: Sexual and Ethical Implications*. Cambridge, MA: MIT Press, 2017.
- Richardson, K. "The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots." *ACM SIGCAS Newsletter, SIGCAS Computers & Society* 45, no. 3 (2015): 290–293.
- Sullins, J. P. "Robots, Love, and Sex: The Ethics of Building a Love Machine." *IEEE Transactions on Affective Computing* 3, no. 4 (2012): 398–409.
- Yeoman, I., and M. Mars. "Robots, Men and Sex Tourism." *Futures* 44, no. 4 (2012): 365–371.
- [Campaign against Sex Robots](#).

Affective Computing

System Manipulation/Nudging/Deception

Issue:
Should affective systems be designed to nudge people for the user's personal benefit and/or for the benefit of someone else?

Background

Emotional manipulation can be defined as an exercise of influence, with the intention to seize control and power at the person's expense. Thaler and Sunstein (2008) call the tactic of subtly modifying behavior a "nudge." Nudging mainly operates through the affective system. Making use of a nudge might be considered appropriate in situations like teaching children, treating drug dependency, healthcare, and when the global community benefits surpass individual benefits. Yet should affective systems be deployed to influence a user's behavior for that person's own good? Nudging can certainly trigger behaviors that worsen human health, but could the tactic be used by affective systems to cue behaviors that improve it? Several applications are possible in health, well-being, education, etc. Yet a nudge could have opposite consequences on different people, with different backgrounds and preferences (White, 2013,

de Quintana Medina and Hermida Justo, 2016). Another key, and potentially more controversial, issue to be addressed is whether an affective system should be designed to nudge a user, and potentially intrude on individual liberty, when doing so may benefit someone else.

Candidate Recommendations

1. Systematic analyzes are needed that examine the ethics of designing affective systems to nudge human beings prior to deployment.
2. We recommend that the user be able to recognize and distinguish between different types of nudges, including ones that seek to promote beneficial social manipulation (e.g., healthy eating) versus others where the aim is psychological manipulation or the exploitation of an imbalance of power (e.g., for commercial purposes).
3. Since nudging alters behavior implicitly, the resulting data on infantilization effects should be collected and analyzed.
4. Nudging in autonomous agents and robots must have an opt-in system policy with explicit consent.
5. Additional protections must be put in place for vulnerable populations, such as children, when informed consent cannot be obtained, or when it may not be a sufficient safeguard.

Affective Computing

6. Nudging systems must be transparent and accountable, implying that data logging is required. This should include recording the user responses when feasible.

Further Resources

The following documents/organizations can be used as additional resources to support the development of ethical affective systems.

- Thaler, R., and C. R. Sunstein. *Nudge: Improving Decision about Health, Wealth and Happiness*, New Haven, CT: Yale University Press, 2008.
- Bovens, L. "The Ethics of Nudge," in *Preference change: Approaches from Philosophy, Economics and Psychology*, edited by T. Grüne-Yanoff and S. O. Hansson, 207–219. Berlin, Germany: Springer.
- de Quintana Medina, J., and P. Hermida Justo. "[Not All Nudges Are Automatic: Freedom of Choice and Informative Nudges.](#)" Working paper presented to the European Consortium for Political Research, Joint Session of Workshops, 2016 Behavioral Change and Public Policy, Pisa, Italy, 2016.
- White, M. D. *The Manipulation of Choice. Ethics and Libertarian Paternalism.* New York: Palgrave Macmillan, 2013.
- Scheutz, M. "[The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?](#)" *IEEE Transactions on Affective Computing* 3, no. 4 (2012): 424–433.
- Grinbaum, A., R. Chatila, L. Devillers, J.-G. Ganascia, C. Tessier, and M. Dauchet.

"[Ethics in Robotics Research: CERNA Recommendations](#)," *IEEE Robotics and Automation Magazine* 24, no. 3 (2017): 139–145.

- "Designing Moral Technologies: Theoretical, Practical, and Ethical Issues" Conference July 10–15, 2016, Monte Verità, Switzerland.

Issue:

Governmental entities often use nudging strategies, for example to promote the performance of charitable acts. But the practice of nudging for the benefit of society, including through the use of affective systems, raises a range of ethical concerns.

Background

A profoundly controversial practice that could be on the horizon is allowing a robot or another affective system to nudge a user for the good of society (Borenstein and Arkin, 2016). For instance, if it is possible that a well-designed robot could effectively encourage humans to perform charitable acts, would it be ethically appropriate for the robot to do so? This design possibility illustrates just one behavioral outcome that a robot could potentially elicit from a user.

Affective Computing

Given the persuasive power that an affective system may have over a user, ethical concerns related to nudging must be examined. This includes the significant potential for misuse.

Candidate Recommendations

As more and more computing devices subtly and overtly influence human behavior, it is important to draw attention to whether it is ethically appropriate to pursue this type of design pathway. There needs to be transparency regarding who the intended beneficiaries are, and whether any form of deception or manipulation is going to be used to accomplish the intended goal.

Further Resources

The following documents/organizations can be used as guides to support the development of ethical affective systems.

- Borenstein, J., and R. Arkin. "[Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being.](#)" *Science and Engineering Ethics* 22, no. 1 (2016): 31–46.
- Borenstein, J., and R. C. Arkin. "[Nudging for Good: Robots and the Ethical Appropriateness of Nurturing Empathy and Charitable Behavior.](#)" *AI and Society* 32, no. 4 (2016): 499–507.

Issue:

A nudging system that does not fully understand the context in which it is operating may lead to unintended consequences.

Background

This kind of system needs to have sophisticated enough technical capabilities for recognizing the context in which it is applying nudging strategies. We could imagine a technical license ("permits") (Omohundro, 2013).

Candidate Recommendation

1. When addressing whether affective systems should be permitted to nudge human beings, user autonomy is a key and essential consideration that must be taken into account.
2. We recommend that when appropriate, an affective system that nudges human beings should have the ability to accurately distinguish between users, including detecting characteristics such as whether the user is an adult or a child.
3. Affective systems with nudging strategies should be carefully evaluated, monitored, and controlled.

Affective Computing

Further Resources

The following documents/organizations can be used as guides to support the development of ethical affective systems.

- Borenstein, J., and R. Arkin. "[Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being.](#)" *Science and Engineering Ethics* 22, no. 1 (2016): 31–46.
- Arkin, R. C., M. Fujita, T. Takagi, and R. Hasegawa. "[An Ethological and Emotional Basis for Human-Robot Interaction.](#)" *Robotics and Autonomous Systems* 42, no. 3–4 (2003): 191–201.
- Omohundro, S. "[Autonomous Technology and the Greater Human Good.](#)" *Journal of Experimental and Theoretical Artificial Intelligence* 26, no. 3 (2014): 303–315.

Issue:
When, if ever, and under which circumstances is deception performed by affective systems acceptable?

Background

Deception is commonplace in everyday human-human interaction. According to Kantian ethics, it is never ethically appropriate to lie, while utilitarian frameworks would indicate that it can be acceptable when it increases overall

happiness. Given the diversity of views on the ethical appropriateness of deception, how should affective systems be designed to behave?

Candidate Recommendations

It is necessary to develop recommendations regarding the acceptability of deception in the design of affective autonomous agents with respect to when and under which circumstances, if any, it is appropriate.

1. In general, deception is acceptable in an affective agent when it is used for the benefit of the person being deceived, not for the agent itself. For example, deception might be necessary in search and rescue operations, elder- or child-care.
2. For deception to be used under any circumstance, a logical and reasonable justification must be provided by the designer, and this rationale must be approved by an external authority.
3. Deception must follow an opt-in strategy and must be transparent to the user, i.e., the context under which the system is allowed to deceive.

Further Resources

- Arkin, R. C., "Robots That Need to Mislead: Biologically-inspired Machine Deception." *IEEE Intelligent Systems* 27, no. 6 (2012): 60–75.
- Shim, J., and R. C. Arkin. "Other-Oriented Robot Deception: How Can a Robot's Deceptive Feedback Help Humans in HRI?"

Affective Computing

Eighth International Conference on Social Robotics (ICSR 2016), Kansas, Mo., November 2016.

- Shim, J., and R. C. Arkin. "The Benefits of Robot Deception in Search and Rescue: Computational Approach for Deceptive Action Selection via Case-based Reasoning." *2015 IEEE International Symposium on*

Safety, Security, and Rescue Robotics (SSRR 2015), West Lafayette, IN, October 2015.

- Shim, J., and R. C. Arkin. "A Taxonomy of Robot Deception and its Benefits in HRI." *Proceedings of IEEE Systems, Man and Cybernetics Conference*, Manchester England, October 2013.

Affective Computing

Systems Supporting Human Potential (Flourishing)

Issue:

Extensive use of artificial intelligence in society may make our organizations more brittle by reducing human autonomy within organizations, and by replacing creative, affective, empathetic components of management chains.

Background

As human workers are replaced by AI, their former employers (e.g., corporations and governments) may find they have eliminated the possibility of employees and customers discovering new equilibria outside the scope of what the organizations' leadership originally foresaw. Even in ordinary, everyday work, a lack of empathy based on shared needs and abilities disadvantages not only the liberty of individuals but also the corporations and governments that exist to serve them, by eliminating opportunities for useful innovation. Collaboration requires sufficient commonality of collaborating intelligences to create empathy – the capacity to model the other's goals based on one's own.

Candidate Recommendations

1. It is important that human workers within an organization have direct interactions with each other, rather than always being intermediated by affective systems (or other technology) which may filter out useful, unexpected communication. Similarly, we recommend human points of contact be available to customers and other organizations.
2. In particular, although there will be many cases where AI is less expensive, more predictable, and easier to control than human employees, we recommend maintaining a core number of human employees at every level of decision-making with clear communication pathways.
3. More generally, management and organizational theory should be extended to consider appropriate use of affective and autonomous systems to enhance their business model and the efficacy of their workforce.

Further Resource

The following document can be used as an additional resource to support the development of ethical affective systems.

Affective Computing

- Bryson, J. J. "Artificial Intelligence and Pro-Social Behavior," in *Collective Agency and Cooperation in Natural and Artificial Systems*, edited by Catrine Misselhorn, 281–306, Springer, 2015.

Issue:

The increased access to personal information about other members of our society, facilitated by artificial intelligence, may alter the human affective experience fundamentally, potentially leading to a severe and possibly rapid loss in individual autonomy.

Background

Theoretical biology tells us that we should expect increased communication – which AI facilitates – to increase group-level investment. This could have the effect of reducing individual autonomy and increasing in its place group-based identities. Candidate examples of this sort of social alteration include:

1. Increased investment in monitoring and controlling children's lives by parents.
2. Decreased willingness to express opinions for fear of surveillance or long-term unexpected consequences.
3. Utilization of "customers" to perform basic corporate business processes such as data entry as a barter for lower prices, resulting also in reduced tax revenues.

The loss of individual autonomy could lead to more fragmented or fragile societies, and (because diversity is associated with creativity) a reduction of innovation. This concern relates to issues of privacy and security, but also to social and legal liability for past expressions.

Candidate Recommendations

1. Organizations, including governments, must put a high value on individuals' privacy and autonomy, including restricting the amount and age of data held on individuals.
2. Educational countermeasures should be taken to encourage individuation and prevent loss of autonomy.

Further Resources

The following documents can be used as additional resources to support the development of ethical affective systems.

- Bryson, J. J. "Artificial Intelligence and Pro-Social Behavior," in *Collective Agency and Cooperation in Natural and Artificial Systems*, edited by Catrine Misselhorn, 281–306, Springer, 2015.
- Cooke, M.. "A Space of One's Own: Autonomy, Privacy, Liberty." *Philosophy & Social Criticism*, 25, no. 1, (1999): 22–53.

Affective Computing

- Roughgarden, J., M. Oishi, and E. Akçay. “Reproductive Social Behavior: Cooperative Games to Replace Sexual Selection.” *Science* 311, no. 5763 (2006): 965–969.

Issue:

A/IS may negatively affect human psychological and emotional well-being in ways not otherwise foreseen.

Background

A/IS has unprecedented access to human culture and human spaces — both physical and intellectual — for something that is not a human. A/IS may communicate via natural language, it may move in humanlike forms, and express humanlike identity. As such, it may affect human well-being in ways not yet anticipated.

Candidate Recommendations

We recommend vigilance and research for identifying situations where A/IS are already affecting human well-being, both positively and negatively. We should look for evidence such as correlations between the increased use of A/IS and any suspected impacts. However, we should not be paranoid nor assume that correlation indicates causation. We recommend robust, ongoing, multidisciplinary research.

Further Resource

The following document can be used as an additional resource to support the development of ethical affective systems.

- Kamewari, K., M. Kato, T. Kanda, H. Ishiguro, and K. Hiraki. “Six-and-a-Half-Month-Old Children Positively Attribute Goals to Human Action and to Humanoid-Robot Motion.” *Cognitive Development* 20, no. 2, (2005): 303–320.

Affective Computing

Systems With Their Own Emotions

Issue:

Synthetic emotions may increase accessibility of AI, but may deceive humans into false identification with AI, leading to overinvestment of time, money, trust, and human emotion.

Background

Deliberately constructed emotions are designed to create empathy between humans and artifacts, which may be useful or even essential for human-AI collaboration. However, this could lead humans to falsely identify with the A/IS, and therefore fail to realize that – unlike in evolved intelligence – synthetic emotions can be compartmentalized and even entirely removed. Potential consequences are over-bonding, guilt, and above all, misplaced trust. Because there is no coherent sense in which designed and engineered AI can be made to suffer, because any such affect, even if possible, could be avoided at the stage of engineering, or reengineered. Consequently, AI cannot be allocated moral agency or responsibility in the senses that have been developed for human sociality.

Candidate Recommendations

1. Commercially marketed AI should not be considered to be a person in a legal sense, nor marketed as a person. Rather its artifactual (authored, designed, and built deliberately) nature should always be made as transparent as possible, at least at point of sale and in available documentation.
2. Some systems will, due to their application, require opaqueness in some contexts (e.g., emotional therapy). Transparency in such instances should not be necessarily during operation, but the systems' working should still be available to inspection by responsible parties.

Further Resources

The following documents can be used as additional resources to support the development of ethical affective systems.

- Arkin, R. C., P. Ulam, and A. R. Wagner. "Moral Decision-making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception," *Proceedings of the IEEE* 100, no. 3 (2012): 571–589.
- Arkin, R., M. Fujita, T. Takagi, and R. Hasegawa. "An Ethological and Emotional Basis for Human-Robot Interaction," *Robotics and Autonomous Systems* 42, no. 3–4 (2003): 191–201.

Affective Computing

- Arkin, R. C.. "Moving Up the Food Chain: Motivation and Emotion in Behavior-based Robots," in *Who Needs Emotions: The Brain Meets the Robot*, edited by J. Fellous and M. Arbib. New York: Oxford University Press, 2005.
- Boden, M., J. Bryson, D. Caldwell, K. et al. "Principles of Robotics: Regulating Robots in the Real World." *Connection Science* 29, no. 2 (2017): 124–129.
- Bryson, J. J., M. E. Diamantis, and T. D. Grant. "Of, For, and By the People: The Legal Lacuna of Synthetic Persons." *Artificial Intelligence & Law* 25, no. 3 (2017): 273–291.
- Novikova, J., and L. Watts. "Towards Artificial Emotions to Assist Social Coordination in HRI." *International Journal of Social Robotics* 7 no. 1, (2015): 77–88.
- Scheutz, M. "The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?" *IEEE Transactions on Affective Computing* 3 (2012): 424–433.
- Sharkey, A., and N. Sharkey. "Children, the Elderly, and Interactive Robots." *IEEE Robotics & Automation Magazine* 18.1 (2011): 32–38.