Code of Ethics on Human Augmentation

S. Mann, Brett Leonard, David Brin, Ana Serrano, Robin Ingle, Ken Nickerson, Caitlin Fisher, Samantha Mathews, R. Janzen, M. A. Ali, K. Yang, D. Braverman, S. Nerkar, K. M.-Sanchez, Zack P. Harris, Zach A. Harris, Jesse Damiani, Edward Button
http://www.eyetap.org/CyborgCode/

Abstract

The possibility that artificially intelligent machines may some day pose a risk is well-known [1].

Less understood, but more immediately pressing, are the risks that *humanistically intelligent* [5, 7] people or organizations pose, whether facilitated by "smart buildings", "smart cities" (a camera in every streetlight), or "cyborgs" with wearable or implantable intelligence. As we augment our bodies and our societies with ever more pervasive and possibly invasive sensing, computation, and communication, there comes a point when we ourselves become these technologies (what Minsky, Kurzweil, and Mann refer to as the "Sensory Singularity"[10]).

This sensory intelligence augmentation technology is already developed enough to be dangerous in the wrong hands, e.g. as a way for a corrupt government or corporation to further augment its power and use it unjustly.

Accordingly we have spent a number of years developing a Code of Ethics on Human Augmentation [9], further developed at IEEE ISTAS 2013 and IEEE GEM 2015 (the "Toronto Code"), resulting in three fundamental "laws".

1 Human Augmentation Code

These three "Laws" represent a philisophical ideal (like the laws of physics, or like Asimov's Laws of Robotics [2], not an enforcement (legal) paradigm:

- 1. (Metaveillance/Sensory-Auditability) Humans have a basic right to know when and how they're being surveilled, monitored, or sensed, whether in the real or virtual world.
- 2. (Equality/Fairness/Justice) Humans must (a) not be forbidden or discouraged from monitoring or sensing people, systems, or entities that are monitoring or sensing them, and (b) have the power to create their own "digital identities" and express themselves (e.g. to document their own lives, or to defend against false accusations), using data about them, whether in the real or virtual world. Humans have a right to defend themselves using information they have collected, and a responsibility not to falsify that information.
- 3a. (Aletheia/Unconcealedness/Technological-Auditability) With few exceptions, humans have an affirmative right to trace, verify, examine, and understand any information that has been recorded about them, and such information shall be provided immediately: **Feedback delayed is feedback denied**. In order to carry out the justice requirement of the Second Law, humans must have a right to access and use of information collected about them. Accordingly, we hold that Subjectrights [6] prevail over Copyright, e.g. the subject of a photograph or video recording enjoys some reasonable access to, and use of it. Similarly, machines that augment the human intellect must be held to the same ethical standard. We accept that old-fashioned, hierarchical institutions (e.g. law enforcement) still have need for occasional asymmetries of veillance, in order to apply accountability to harmful or dangerous forces, on our behalf. However such institutions must bear an ongoing and perpetual burden of proof that their functions and services justify secrecy of anything more than minimal duration or scope. Application of accountability upon such elites - even through renewably trusted surrogates, must be paramount, and a trend toward ever-increasing openness not thwarted.
- 3b. Humans must not design machines of malice. Moreover, all human augmentation technologies shall be developed and used in a spirit of truth, openness, and unconcealedness, providing comprehensibility through immediate feedback. (Again, feedback delayed is feedback denied.) Unconcealedness must also apply to a system's internal state, i.e. system designers shall design for immediate feedback, minimal latency, and take reasonable precautions to protect users from the negative effects (e.g. nausea and neural pathway overshoot formation) of delayed feedback.
- 3c. Systems of artificial intelligence and of human augmentation shall be produced as openly as possible and with diversity of implementation, so that mistakes and/or unsavory effects can be caught, not only by other humans but also by diversely competive and reciprocally critical AI (Artificial Intelligence) and HI (Humanistic Intelligence).

A metalaw states that the Code itself will be created in an open and transparent manner, i.e. with instant feedback and not written in secret. In this meta-ethics (ethics of ethics) spirit, continual rough drafts were posted (e.g. on social media such as Twitter #HA-Code), and members of the community were invited to give their input and even become co-authors.

2 The Second Law

The First Law is well-documented in existing literature on metasensing, metaveillance [8], and veillametrics [4]. Interestingly, the City of Hamilton, Ontario, Canada, has passed the following bylaw, relevant to the First Law of Human Augmentation:

> "No person shall: Apply, use, cause, permit or maintain ... the use of visual surveillance equipment where the exterior lenses are obstructed from view or which are employed so as to prevent observation of the direction in which they are aimed." [3].

The Second Law asserts that systems that watch us, while forbidding us from watching them, are unfair and often unjust.

2.1 The Veillance Divide is Justice Denied

In the new, "transhumanistic era", some machines will acquire human qualities such as AI (Artificial Intelligence), and some humans will acquire machine-like qualities such as near-perfect sensory and memory capabilities. Irrefutable recorded memories - suitable as evidence, not mere testimony - will challenge many of our old ways, calling for updated ethics that serve the interests of all parties, not just those with power or authority. Our greatest danger may be a "(sur)Veillance Divide" where things (Internet of Things) and elites may record with perfect memory, while ordinary people are forbidden from seeing or remembering. Therefore, we propose the following pledge, to clarify the need for fairness, equality, and two-way transparency:

- 2a(i). I pledge to not surveill or record any individual or group while simultaneously forbidding that individual or group from recording or sousveilling me.
- 2a(ii). I pledge to respect the needs of others for the sanctity of their personal space. I will negotiate any disagreements reasonably and with good will.
- 2a(iii). If I witness a crime against fellow humans, whether perpetrated by low-level criminals or by elites or by authorities, I will aim to record the event, overtly or covertly (whichever is appropriate). I will aim to make such recordings available to injured parties.
- 2a(iv). I will maintain that, with few exceptions, being surveilled while simultaneously being forbidden from sousveilling, is itself an injury. Therefore, if I witness any party being recorded, while that party is simultaneously prevented from recording, I will aim to record the incident, and to make the recording available to the injured party.
- 2a(v). I will make a best effort to be informed of escrow storage (e.g. "videscrow"), so that when recording others, there can be "temporary exclusions" on retroactive recording until disagreements may be adjudicated. Here the burden-of-proof is on the party prohibiting unescrowed recording.
- 2a(vi). I will try not to be provocative or confrontational, assuming the worst about others. But the light that I shine and the recordings I take may thwart injustice. It is possible to apologize and make amends for too much light. Too little can be lethal.

3 Conclusion

We take here an imporant first step toward the Human Augmentation Code 1.0. This is a "living document" and we are open to contributions from all, as it evolves.

References

[1] N. Bostrom. Ethical issues in advanced artificial intelligence. *Science Fiction and Philosophy: From Time Travel to Superintelligence*, pages 277–284, 2003.

[2] R. Clarke. Asimov's laws of robotics: implications for information technology-part i. *Computer*, 26(12):53–61, 1993.

[3] M. Fred Eisenberger and C. C. Rose Caterini. City of hamilton by-law no. 10-122. May 26, 2010.

[4] R. Janzen and S. Mann. Sensory flux from the eye: Biological sensing-of-sensing (veillametrics) for 3d augmented-reality environments. In *IEEE GEM 2015*, pages 1–9.

[5] S. Mann. Humanistic intelligence/humanistic computing: 'wearcomp' as a new framework for intelligent signal processing. *Proceedings of the IEEE*, 86(11):2123–2151+cover, Nov 1998.

[6] S. Mann. Computer architectures for personal space: Forms-based reasoning in the domain of humanistic intelligence. *First Monday*, 6(8), 2001.

[7] S. Mann. Wearable computing: Toward humanistic intelligence. *IEEE Intelligent Systems*, 16(3):10–15, May/June 2001.

[8] S. Mann. The sightfield: Visualizing computer vision, and seeing its capacity to" see". In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 618–623. IEEE, 2014.

[9] S. Mann. Keynote address: Code of ethics for the cyborg transhumanist era. In *Second annual conference of the World Transhumanism Association*. http://www.transhumanism.org/tv/2004/, August 5-8, 2004.

[10] M. Minsky, R. Kurzweil, and S. Mann. The society of intelligent veillance. In *IEEE ISTAS 2013*.