Title of the Ethics Paper Goes here

Introduction

This article proposes an ethics of “Human Computer Interaction”. We bring together philosophical sources and references to ethical thinking within HCI. In the process, we contextualise ongoing discipline-specific work such as the efforts of the ACM SIGCHI Research Ethics Committee — and broaden the scope of that work to reflect on human interaction with computers tout court.

Our project stems from a documented need to further develop the existing understanding of the pro-social use of computational technologies. “Interaction” becomes a central theme in this work insofar as the perspective that we develop hinges on various forms and conceptions of interaction with and within information systems.

Our reason for taking such a theoretically-broad stance is the awareness that humanity has entered into a new historical era, variously referred to as the anthropocene, the information age, the fourth Industrial Revolution, and, indeed, the dawn of the Novacene (Lovelock). Many of the historically-novel and existentially-salient aspects of life in the early 21st Century have a legacy going back to the industrial revolution and the Enlightenment. Others date further back. Nevertheless, the theoretical structures left to us by the Enlightenment age are as ill-adapted to our current concerns as the steam engine and the guillotine.

To ameliorate the harms and capitalise on the benefits of our present post-modern condition, we need a way to understand human-computer interaction ‘at scale’. The philosophy of technology will run throughout our work here, especially as regards our methods. We offer a brief recapitulation of key points from this tradition in the section on Related Work. Emphatically, in this paper we are applying the philosophy of technology and have no pretentions to make original contribution to that field.

The paper as a whole focuses on the following questions:

1. How are notions of ‘ethics’ used and applied in contemporary information systems?
2. How do the systems we engage with ‘react back’ on our ways of thinking?
3. How do computational systems apply or enact the ethics that we apply, or indeed develop ethics of their own?
4. And, lastly: what is the overall narrative or genre in which these questions can be discussed and pursued further?

Further notes:

- How do we use ethics
- Motivation

Method

We survey references to ethics and philosophy within computing literature — along with references to technology, machinery, and computing within philosophy. Based on aligning this material we propose first a taxanomy and then a set of ethical guidelines.

Our aim is not completeness but robustness. On the one hand we attempt to account for the pervasive use of information and communication technology (ICT), and the potential future development of artificial intelligence.

At the outset, our initial framework-for-the-framework recapitulates the division mentioned in the Introduction, which mirrors the major sections of the article. We consider:

1. The interfaces in which ethics can be applied (species and robots).
2. The change in behavior that comes from the ’benefit’ to make people interconnected as a global mind.
3. Until what point social interaction can change and exist?
4. We reference and on existing surveys, including “Machine Implementations of Ethics” and “Peer Production: A Modality of Collective Intelligence”.

Scope

Corresponding to the broad aims outlined above, alongside HCI, Human-Computer Cocreativity (HCCC) and Computational Social Creativity (CSC) are inspiring domains. In HCCC, creativity is attributed to collectives; CSC embeds models of social creativity within computational systems. Our aim will be to look in more detail at what what humans co-create with computational media. While our focus is on modern computing machinery, in order to cast a wide

net we think about computers as ‘effective systems’ or even just as tools or machinery. Tools and language have been with us since the dawn of humankind.

1.0 Philosophical foundation for a contemporary ethical practice

> How are notions of ‘ethics’ used and applied to contemporary information systems?
>
> The interfaces in which ethics can be applied (species and robots).

To address this question properly we need to understand the key concepts.

1.1 Western philosophy and ethics definition

Human beings have always been interested in categorizing their behaviour. Classification and criticism, personal and interpersonal, about actions taken towards the world. When this criticism is made, there is no way that it will not be realized subjectively (lens argument), even if it embraces generalization. This recursive feedback to outselves has allowed the analysis and assessment of actions that we deem meaningful - or that others consider significant in us - originating fundamental positions for human understanding from the human. It is not in vain that ethics derives from the Greek ”ēthikós” (ἠθικός) that means ”relating to one’s character” - as a relational loop. Indeed, a deeper etymology points to both habit and environment; ἤθεα meaning ”accustomed place” (as in ἤθεα ἵππων — ”the habitat of horses”, Iliad, 6.51115.265). This points to an ”ethological” side of ”ethics”, which is pursued, e.g., by Spinoza (as per Deleuze’s reading). Importantly, our environments include others and provide grounds for interaction. Just as I consider myself, I can imagine the way others consider me, consider them, and develop collaborative processes for all this. Right here comes ethics and the way it will happen in society. Paul and Elder (2003) define ethics as: a nondetermined set of concepts, principles and metarules that guide us in determining what behavior (acting towards) helps and/or harms sentient creatures. Taking this as our working definition of ‘ethics’ allows us to develop ethics helps to create a relationship structure.

Further notes:

- 1.A

- Phenomenology and being
- Ethics and Ethology in Homer

1.1.1 Responses in computing literature

These days, given the transmutation of information and increased reception of things around us, technology can even help us to revive and rethink the way these thoughts affect us culturally. For example the work of Kantosalo and Schneiderman: using computers to help us ”think” about creativity. (Anna Kantosalo and Ben Schneiderman (using computers to think about ‘creativity’))

1.2 Holistic views of philosophy

Philosophy is inherent in the complexity and uniqueness of each culture. When applied from a broad point of view and not only focused on the human, it can conquer and integrate as its baggage other types of entities such as artificial intelligences or non-human organisms. As is the example of panpsychism (Seager 2006), where it is common to have a naturalistic account of the world, for only from the point of view of some such account can the issue of mind’s place within the natural world arise. Even the philosophy of mind has in itself a position. Not only as a philosophy, but as a starting point for philosophy to begin. There is a beginning that points to a vision in which all the positions taken before that thought make sense. When using positions that are global in terms of beings that exist, an ethical framework forces us to consider computers as creatures.

Further notes:

- Holistic views of philosophy

1.2.1 Responses in the computing literature

Given the holistic point of view, which in itself is linked to space and context - the reflections made today often map the past. The way we relate - to each other, to artificial elements and the environment - is what has been driving us, we might want not only to map but also revisit past.

2.0 Embodied cognition, social intelligence, collective intelligence

> How do the systems we engage with ‘react back’ on our ways of thinking?

> The change in behavior that comes from the 'benefit' to make people interconnected as a global mind.

The survey of the philosophy in the previous sections gave a general outline of the role of the being in nature. Here, we focus on this being's knowledge, and how this being — he, she, or it — can inform the development of 'thinking machines', and to what extent some of these terms are intrinsically social.

Certain landmarks in the evolution of contemporary perspectives on these themes include:

1. Kant's take on about interpersonal relationship;
2. The way intrapersonal evolution is seen through Freud's eyes;
3. The extent to which we should and can be analytical in developing an ethical stance towards 'the whole', with Carl Jung;
4. Ongoing developments of with tools like the Helmholtz Machine and active inference.

The fields of cognitive science and philosophy express one others' limits. In the capsule history outlined in points 1-4 above, the work of the thinkers mentioned increasingly intersect with science. In the process their work begins to articulate what 'mind' means in practical terms: what it means to have one; how can we catalog various species of minds for the future; how these interact; and the various points of view on 'mind' that exist (functionalists, patternists, etc.). We can use this material as starting point to articulate our relationship with the taxonomy of ethical perspectives.

Responses in the computing literature

In contemporary computing:

1. the activities which are referred as creative are typically seen as social and socially computational (mention Anna's paper here)
2. virtual distributed agency and behavior is exactly what is happening in the physical world
3. current approaches draw on, e.g., bioinformatics and computational neuroscience
4. the theory of the Cyborg manifestos as an ethical proposal (Donna Haraway)
5. current work on computers and interaction.

AI and interaction with computers more generally is understood as a potential force for "good" — if that is understood as pro-social and evolutionary — and also as a source of risks.

3.0 Reprise: Evolution regarding all of these

> How do computational systems apply or enact the ethics that we apply, or indeed develop ethics of their own?
>
> Until what point social interaction can change and exist?

This section repeats this reflective movement from Section 2.0, through the sphere of artificial intelligence (AI). When we think about ethics and AI, what is lost in translation? What could potentially change for the better? In order to address these questions we take a running leap, building momentum with thinking about evolution more broadly.

Histories of the evolution of intelligence (sociality & tools being key focal points). Theories of evolution, e.g., Baldwin (and later derived work by Hinton and others). Derrida's concept of différance.

Based on the points raised as discussion in the previous sections mention until what point evolution plays a or the major role. How future AIs will encompass some of the evolutionary paradigms we faced and how our ethics project will not be ruined in future decades - getting to the point where evolution might be quicker virtually (as a type of evolution).

Further notes:

- Language is mapping thinking
- Evolution

Responses in the computing literature

The mapping of evolutionary techniques and parallel thinking (social behavior also mapped and check if this doesnt exist elsewhere). Metacognition as assessment and metamemory as understanding if we remember is true and the access we can have. Cognitive psychology approaches to AI (maybe connect this to reinforcement learning and behavior?) Current approaches to model ethics in computers as values and the ones that model only the environment that will give rise to the values in the first place (2021 literature): Predictive Processing and Active Inference (bring embodiment to the discussion here); if "Ethical AI" is important or a more globalist perspective: Notice that now that computers are involved, the way we think about ethics and so on is likely to change.

4.0 Narratives, genres, and disciplines: How do we talk about HCI ethics?

> What is the overall narrative or genre in which these questions can be discussed and pursued further?
>
> We reference and on existing surveys, including “Machine Implementations of Ethics” and “Peer Production: A Modality of Collective Intelligence”.

E.g., are these considretainos in fact proper to Philosophy, after all? Or are these themes that can be addressed within Computing? Or should we refer to Law? Or Religion? Or Science Fiction? Or Art? Or something else?

Responses in the computing literature

We see how mediatic engagement with ethics is scaffolded by directing attention to: (1) Ethical impact agents; (2) Implicit ethical agents; (3) Explicit ethical agents; (4) Fully ethical agents; (5) How are ethics really used in systems - that in itself.

Ethics Taxonomy

An ethics taxonomy is presented as a mapping of values and positions we and machines can take now and in the future regarding the questions raised such as: 1) how can we and machines establish a true and positive relationship with each other in points such as 1.1) designing other machines or (artificial) humans; 1.2) impact other elements of the society; 1.3) change ourselves; 2) what does it mean to be ethical towards something using an abstract definition; 2.1) what being means comes from above; 2.2) towards something also comes from above; 2.3) abstract definition comes from language also from above; 3) define and utilize this taxonomy based on interaction, social behavior, design and engineering, be computing␣platform-agnostic and topic-agnostic, and how machine ethics is right or wrong as a separare domain, how to imply ethics works and doesn’t work; propose meta-ethics guidelines on how can we create ethical guidelines that create ethics.

Discussion

Have we learned anything that’s relevant for practice? Maybe here is a good time to return to some of the debates that look at ”cre-

ativity" in a more mainstream sense, e.g., Anna Kantosalo and Ben Schneiderman about creative systems and social inclusion vs exclusion? From the point of view of "Methods", hopefully we will have clarified at the start why we think this sort of activity could lead to new insights! We will build a thought experiment in the text to utilize the raised taxonomy.

As related work we should specifically engage with Floridi:

> With distributed agency comes distributed responsibility. Existing ethical frameworks address individual, human responsibility, with the goal of allocating punishment or reward based on the actions and intentions of an individual. They were not developed to deal with distributed responsibility.

This is clearly germane, and we can go further with reference to "systems with emergent properties"; so, if distributed agents produce e.g., environmental degradation, that's not "ethical", and the system as a whole "should" find ways to improve its behaviour. This sort of thing is thought about in Elinor Ostrom's economics. A particular concern of Taddeo & Floridi here seems to be "autonomy" of AI, and "self-determination" of humans. But in the case of HCI/HCCC it's not totally clear that either of these criteria apply. In HCCC it's much closer to anthropotechnics.

Hopefully we can provide some new insights here.

Further notes:

- Case studies reprise

Related work

Alongside philosophers of technology mentioned in the Introduction, we can point to more popularly-oriented books such as ("Creativity and Ethics", "Technology and the virtues: A philosophical guide to a future worth wanting", "Made by Humans", "Machines that Think", "How AI can be a force for good" — and connect all these topics with political, scientific and visionary points that authors made in time.

Further notes:

- Philosophy of technology
- https://sigchi.org/ethics-committee/

Conclusions and Future work

In addition to the questions in the introduction, as a result of the theoretical work developed here we sohuld be able to offer at least tentative answers to the following questions: 1) How can I practically engage with these issues as a computer science researcher?; 2) What are future steps and possibilities to research ethics, to practice ethics and relate this to other ethics roles (as we did in all the text) (e.g maybe also at the governmental level; 3) How do interfaces and other concrete-relationships-between-people-and-things embodied behavior and its limits for ethics (where our theory becomes virtual and link to haraway); 4) How do I relate to knowledge, what it means to know or to cognise; with/to the whole body of historical philosophy, science, inquiry, and maybe AI and tech systems?

If nothing else this should be seen as an alternative to ”Ethical AI” as it is currently practiced (either as governance of real-world systems or imagining the future). By focusing on interaction we mean to develop a route to ongoing improvementment to HCI ethics overall (in an eternal golden braid!).

Further notes:

- Conclusion