IRIE
International Review of Information Ethics Vol. 31 (08/2022)

Marc M Anderson and Karën Fort

Human Where? A New Scale Defining Human Involvement in Technology Communities from an Ethical Standpoint

Abstract:

We undertake an expansive examination of the terms Human-in-the-loop, Human-on-the-loop, Human-out-of-the-loop, and Human-in-command, as used recently in AI development, relative to their ethical implications and implicit assumptions. Tracing the history and development of the 'Human …' terms, we explore the contexts and uses present from their beginnings. We follow with a discussion of the ethical outlook which the origins of the terms and their recent rebranding for AI development under the notion of *oversight* have engendered. Drawing upon certain insights of Bruno Latour for support, we suggest that Latour's 'forgotten ethical intermediaries', folded into our technologies, have their analogue in the view of the human as a component of automated systems alternating with a role of human oversight. We argue that a more ethical human relation to technology can be recovered through an expansive emphasis on human *participation* in technology producing communities. Finally, we present a flexible new scale, the IGP scale, to rate such participation.

Keywords: ethics, artificial intelligence, human in the loop

Agenda:

Authors:

Dr Marc M Anderson:

- LORIA, UMR 7503, Université de Lorraine, Inria and CNRS, Campus Scientifique, 615 Rue du Jardin-Botanique, 54506 Vandœuvre-lès-Nancy
- marc.anderson@inria.fr ; dr.marcanderson@gmail.com

Dr Karën Fort, Associate Professor:

- Sorbonne Université/LORIA, UMR 7503, Université de Lorraine, Inria and CNRS, Campus Scientifique, 615 Rue du Jardin-Botanique, 54506 Vandœuvre-lès-Nancy
- karen.fort@loria.fr

Funding:

The research leading to this article was funded through the AI-PROFICIENT project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 957391.

1. Introduction

With the widespread use of artificial intelligence (AI), related ethical issues are more and more discussed and taken into account. However, the human interaction with AI is defined using terms, such as *Human in the Loop*, *Human on the Loop*, and *Human out of the Loop,* which were designed for other purposes. We argue that there is an original tension inherent in these terms, which comes from the notions of human as component and human as overseer. The 'Human…' terms were not originally ethical in intent having been originally created for control system engineering and later ML, among others. It would be difficult and unwieldy to modify them to serve an ethical intent, particularly a practical ethical intent, such as is needed in AI ethics, and the contemporary attempt to use them for such a purpose is confusing and counterproductive. In this article we will consider the development and efficacy of the terms *Human in the Loop*, *Human on the Loop*, and *Human out of the Loop*, as well as the more recent *Human in Command*. We begin by surveying their origin and development. From there we consider the variety and sometimes confusion of meanings that the concepts have been used for. Drawing on some suggestions by Bruno Latour regarding technology, we move a step beyond Latour and propose an expansion of the human role in relation to technological creations.

2. History and Development of the 'Human …' Terms

2.1 Human in the Loop (HITL)

The beginning of our interest in the 'Human …' terms was sparked in the context of applied AI ethics research in which we had to interact closely with technology developers and industrial partners. There we found the term Human in the Loop and related terms being tossed about fairly indiscriminately. This led, within our own AI ethics team, to questioning and discussing what the terms really meant, what they could mean ethically, and ultimately to attempting to track down the origins of the terms.

Our subsequent research on HITL and associated terms was carried out manually using multiple searches for multiple combinations of words in the terms, e.g. 'human in the loop', 'in the loop', 'the loop', etc., within Internet Archive primarily and Google Scholar secondarily. Use of the former is laborious but particularly helpful as it gives access to older documents and types of documents which do not fall within standard journal publication parameters. Based on that research we are confident that the progenitor of all the more recent 'Human …' terms is Human in the Loop which began to be used in publications in the 1950s. Joseph Shea's summary of the related terms, and closely related instances which do not discuss humans, e.g. "components in the loop" (Brown 1948), suggests that the term began to be used sometime in the 1940s in informal spoken communication relative to control loops.

The first distinct published mention of the term HITL that we were able to find occurred in "Problem Notes" of the *Report of Naval Progress* of 1958, prepared by the Naval Research Laboratory (Birmingham et al., "Target Tracking" 20).[1] It seems likely however that the exact term or something similar was in use as far back as the mid-1940s.

All of the early explicit or implicit uses of the term are in the context of control systems, and implicitly gathered around solutions to problems of precision and stability in missile fire-control systems, e.g. Birmingham and Taylor in 1954 (1749). The major concern in early references appears to be the variability of human behaviour, or as Birmingham puts it, "the human operator modifies his transfer function and alters his gains to suit the control task with which he is confronted" (1752). In other words, the human can alter behaviour to compensate for design flaws in the system. "This adaptability on the part of the man is, of course, a great boon to the control designer, since he can rely upon the human to make the most of any control system, no matter how

[1] A less exact mention: "Some experimentation on the *tie-in of the human operator to the control loop* of an airborne navigational digital computer system," [italics ours] occurs in Bennett in 1957 (68).

inadequate . . . Yet, this very adjustability renders any specific mathematical expression describing human behavior in one particular control loop quite invalid for another man-machine arrangement" (1752). Thus, this human adaptability is undesirable. Instead, the optimum performance of control systems is achieved when "*the* [behaviour] *required of the man is, mathematically, always as simple as possible, and, wherever practicable, no more complex than that of a simple amplifier*" (1752). Ultimately, Birmingham's view is that humans should be replaced by mechanical components whenever feasible but should replace mechanical components in turn, whenever the humans are *also* necessary as monitors of the machine for safety reasons, *if they are more efficient than the components in question* (1753).

Thus, the *human in the loop* as control system component is present from the beginning together with the concept of the human as merely monitoring the system for safety reasons (Bennett 68), and retains this monitoring meaning as it spreads to more complex computer systems, e.g. Stotz in 1963 (328). But for Birmingham and Taylor at least, the monitoring aspect is completely separable from the aspect of the human as a 'component within the system,' and it is the *latter* which they mean by *human in the loop*. Birmingham and Taylor are referenced widely by those who come soon after, e.g. (McRuer and Krendel 403) and their outlook is both intimately linked to military research[2] and taken up concurrently by military engineering developments.

The source of the initial concept and term HITL is thus the psychological study of humans in control systems from an engineering perspective, in order to solve problems of precision encountered in warfare and its near 'cousins': flight, navigation, etc. HITL begins as *a term defining how to make optimal use of the human as a mathematizable component of a simple mechanical/electrical control.*

Over time HITL has taken on multiple meanings beyond the original and its usage is confusing. In 1976 (Gschwind 16) is using the term in its original sense. (Caid and Simmon 20) in 1979 use the term in the sense of a human as a complex component making choices regarding parameters in a system designed to convert graphs and numeric tables to equations, where the computer system cannot yet perform the human actions efficiently enough. In 1985 (Wixon et al. 145), the term is again used in its original sense, but with regard to an operator acting as a hidden component of an automated system in the development of a User Derived Interface. The term is used by (Reynolds 206), 1988, to describe simulations which involve humans acting in real time. Later similar uses are most often linked to military applications, e.g. (Hopkinson and Sepulveda 1250), 1995, where the human is being trained by the simulation, and viewed as a trainable component of a larger system for military command and control purposes, somewhat similar to the 1950s use.

Steiber (92-93) on robotic control in the Space Station, 1998, uses the term in the sense of the human being aided in a task by motion control algorithms, reserving 'Automatic control' for human monitoring of fully automated operations. Sometimes HITL is used as shorthand for any type of human involvement at all – 'simply being involved' – in a system e.g. (Loy 31), 1993, and (Rissland and Daniels 60) in 1995. By the early 1990s it begins being used regularly in Machine Learning, e.g. (Ayuso 345), in 1993, to mean human training of an algorithm.

2.2 Human on the Loop (HOTL)[3]

The earliest published reference of the term HOTL we found was in (McLaughlin et al.) in 2000, which notes the challenges humans present when included within control loops incorporating haptic devices, while discussing algorithmic solutions to stabilizing the control loop. Here the use echoes its earliest uses with regard

[2] Birmingham worked for the Naval Research Laboratory on multiple projects.

[3] Replaced by HOL in some publications. Confusingly this term is also replaced by *human-over-the-loop* in some publications, e.g. PDPC. 2020. Model Artificial Intelligence Governance Framework (2nd edition).

to *human unpredictability and the difficulty of mathematizing human behaviour*. The human is an unavoidable *component* – the haptic devices are built for human use – which supervisory adaptive control is compensating for. In 2011, Stouch et al. discuss the use of evolutionary algorithms for military drone air mission planning (1696). Here the human is *monitoring* ongoing drone mission planning progress. However (Cummings et al. 1) use the term with regard to humans *actively guiding* automated planners being used to generate plans for unmanned vehicle search, track, and destroy missions.

In its earlier use by McLaughlin, the '*on*' of HOTL seems to derive from the relation of the human to the virtual museum objects which are the focus of haptic exploration. The haptic devices allow a "hands-on" engagement of what is *on* the surface of the object. This is not a monitoring of the control loop in any way, nor a guiding of a process. The original conceptual source of the term HOTL seems to be *an attempt to express a relation in which the human accesses something 'through' an automated system*. This original use is then lost later, at least in part.

2.3 Human Out of the Loop (HOOTL)[4]

The earliest published reference we found to HOOTL appears in 1963 (Shea 7). Shea was manager of the Apollo program. He refers to "man out of the loop," as one among many 'catch phrases' already in use. He discusses the issue in terms of what tasks a man can do best in contrast with what the automated system can do best, and hesuggests that HOOTL often comes down to a question of semantics. In 1966 (Alper 95) we find the notion that a man can be "mechanized out of the loop," supporting Birmingham's earlier view that humans should be replaced by mechanical components whenever they can be. (Sanders 62), in 1988, describes HOOTL as a 'military jargon' for removing human presence entirely from a military control system. How old the term is as military jargon is unclear, but it probably predates the 1960s given Shea's 'catchphrase' description and the history of close cooperation between NASA and the military (Day). The term is used similarly by (Batali 18), in 1995, as removing the human, in the context of the AI developments of that period. Batali denies that it is possible. Porathe and Prison (2864), 2008, note the problem of loss of situational awareness of humans retaking control from a fully automated 'out of the loop' situation, in ship navigation, indicating that HOOTL does not preclude monitoring by the human.

A number of authors view HOOTL as poorly defined, e.g. Merat et al., in 2018. They attempt to define HOOTL in the context of automated vehicles as "not in physical control of the vehicle, and not monitoring the driving situation, OR in physical control of the vehicle but not monitoring the driving situation" (88). Adding ambiguity however, they extend monitoring to both the autonomous system and the car's surroundings, while characterizing the 'Human …" terms as non-discrete. (Hammes), in 2020, notes a contemporary attempt to define HOOTL relative to automated weapons, as: capable of operating without human input.[5] He suggests that such definitions are inadequate, since human input in designing, building, positioning and programming such weapons will always be needed, until artificial intelligence becomes capable of taking over all these tasks.

HOOTL is thus only very vaguely defined from the beginning, taking its start generally from *the idea of removing the human from an automated system*, but varying greatly in whether that can be accomplished and to what degree, and whether the human can get back into the system but still be considered HOOTL. In addition, the term seems to arise at least partially in response to military situations where a human might be excluded from 'momentous decisions,' e.g. the launch of nuclear missiles (Sanders), or targeted killing by UAVs without human input (Wagner 3), in 2011.

[4] Replaced by OOTL in some publications.

[5] (Docherty et al. 2012).

2.4 Human in Command (HIC)

HIC is the most recent of the terms to be formulated. Aside from voluminous direct military usage describing human to human military command, it appears to have been coined or borrowed in 2019 by the European Commission's High-Level Expert Group on Artificial Intelligence (hereafter HLEG) *Ethics Guidelines for Trustworthy AI*.[6] Later uses reference the HLEG guidelines (Fanni et al.). It is defined as: "the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation" (HLEG 16).

The HIC term appears to extend the notion of oversight, one of the original but secondary meanings of the 'Human …' terms to a very high level of generality, essentially suggesting that *all potential human-AI relations can be brought under the umbrella of oversight*.

We have summarized the various uses in Table 1. below.

Table 1. The Multiple Meanings of 'Human in the Loop' terms

Term	Human Status	Publication Year(s)[7]
Human in the Loop	Component of a control system to be simplified and replaced when possible	1954; 1959; 1976
	Monitor of a control system	1954; 1963
	Complex component of a calculating system to be replaced when possible	1979
	Hidden complex component of an automated system	1985
	Term for *a simulation* which includes real-time human action[8]	1988
	A trainable component of a simulation	1993; 1995
	Being aided by motion control algorithms (robotics)	1998
	Being involved in any way *at all* in the use of a computerized system	1993; 1995
	Training an algorithm (Machine Learning)	1993

[6] We have found several online non-published references as early as 2017, but still in contexts related to the development of the HLEG.

[7] Publication year of first published reference in which the authors found the related meaning (plus additional instances of similar meaning)

[8] Focus is on the *simulation*

	Capable of intervening in every decision cycle of a system (HLEG)	2019
Human on the Loop	Unavoidable component which a control system is compensating for	2000
	Monitoring the progress – not the algorithm or errors – of an algorithm driven process	2011
	Actively guiding an automated process	2012
	Monitoring, and intervening in the design cycle of a system (HLEG)	2019
Human out of the Loop	Catch phrase for removed generally from a control system	1963
	Replaced by a mechanism in a control system	1966
	Removed entirely from a control system	1988
	Removed to some degree from an automated system	1995; 2020
	Removed from an automated system but able to retake control	2008
	Removed from immediate input in 'momentous decisions,' e.g. targeted killing by UAVs	2011
	In physical control/or not (of vehicle) but not monitoring (vehicle or surroundings) + a non-discrete state	2018
	Capable of operating without human input	2020
Human in Command	Capable of overseeing all aspects of an AI system (HLEG)	2019

3. Discussion of the Terms Relative to Ethics

Given the variety of uses of the 'Human…' terms, the implications of the original uses of HITL, the differences in what the terms refer to and the degree of intent in using the terms, it seems to us that an ethical consideration of these terms is well overdue. Such an ethical consideration should address attempts to salvage the terms by simply adding related new terms, the attempt at generalizing around the oversight notion, and above all the logical inconsistency inherent in the original intent behind the terms.

3.1 The Proliferation of New Terms

Our survey of the meanings of the terms is limited, but comparing those meanings shows that the terms are inadequate for accommodating various ethical human-'machine' relations. For example, the HLEG definition of HITL, abandons a number of the many meanings which the term has been used for: the human as an undesirable/unpredictable component in a control system, the human being trained by interaction with an AI embedded simulation, human involvement *as such* in automated systems, and active human training of an algorithm – the ML usage – rather than simply being in potential control of it in its local consequences. HOTL, meanwhile (HLEG definition) takes on an original, but secondary, meaning of HITL – monitoring – and adds human oversight of system design, but leaves out monitoring of aspects of the larger process (results monitoring), active guidance, and the human as component. HOOTL while derived from the intent to remove

the human from an automated system, vacillates between the long-term intent of *removing humans entirely from such systems*, e.g. (Shea) and (Hammes), and the short-term intent of *removing a human from this particular system*, e.g. (Batali) and (Merat et al.). It is also ambiguous as to degree, e.g. (Porathe and Prison) and (Batali). How far 'out' does the human have to be to be able to say that a control system is not being monitored? Finally, HIC, is a very new term, whose practicality has yet to be tested.

The terms have been used at relatively the same time with different intent and meaning. Moreover, their meanings have not progressively evolved over time to satisfy the new contexts of technological development. Apart from Batali, we find little or no *disagreement* relative to the use of the terms. New terms are formulated for different contexts.[9] This indicates deep uncertainty in the terms. We need clear terms to think clearly and develop ethics properly however.

3.2 Generalization

The HLEG definitions generalize the human-machine relation in terms of degrees of separation of the human from the automated system. Human and system are viewed as distinct. The human controls the system from outside. The generalization proceeds in layers: HITL - the human can manipulate or 'steer' the system, HOTL – the human can 'steer' and watch the system, HIC – the human can 'steer', watch, switch the system off, decide if it will be built, etc. This omits the ethically problematic *component* characterization of the human. But ignoring the component problem merely avoids ethical discourse regarding practical ethical difficulties which are ongoing, namely *can a human be ethically viewed as part of an automated system at all*? Conversely *can a human be entirely separated ethically from technology which humans have created*?

If generalization were sufficient, then human relations to contemporary technologies would be no more ethically problematic than early humans paddling on a log, where on, off, and controlling the log, or not, might have sufficed. Generalizing the terms may be sufficient for an engineering view, but not for an ethical engagement. Human technology has advanced considerably. We are now explicitly *in* our technology, and it is *in* us, in complex and subtle ways.

Broad generalization, as in the HIC term, draws so much into it that it avoids the problem of the human relation to AI (and technology) as such. HIC potentially covers every problem that might arise with the AI. Yet, only with the greatest difficulty can this be practically implemented in complex technological contexts so as to "design this system according to HIC." Seemingly simplified oversight aspects, e.g. capability to shut down the AI, fail in complex contexts. With an AI deeply integrated into a production line in manufacturing, for example, you cannot approach every problem which arises by saying "shut off the AI." Adopting this approach gives a false sense of security regarding ethical issues.

3.3 Logical Inconsistency with Ethical Consequences

It seems to us that the 'Human …' terms, taken as a group, have arisen from an unquestioning acceptance of two main perspectives which are polar opposites: the human as component, and the human as overseer. This may be because technology advances in AI and control systems appear to eat away at 'unique' human capabilities, by gradually taking over tasks that only humans could do. If Birmingham, and (De Lemos 165-66), are right in suggesting that the goal of engineering is to remove humans entirely from automated systems, could it be that the 'Human…' terms merely function as shorthand and often ambiguous placeholders for measuring the degree to which humans are removed?

If so, then by way of their origins, which bring mainly *negative conceptions of human participation* to the fore, the 'Human …' terms are inadequate for ethical tasks of viewing AI and control systems from a human centered

[9] Besides Shea's 'across the loop,' we have found *human-in-control*, *human outside the loop*, *human above the loop*, *human over the loop*, *human under the loop*, and *meaningful human control*, among others.

ethical perspective. Humans, in the 'Human …' terms, *are being measured by how dispensable they are*, but this is logically, inconsistent with *an ethics for humans*. Viewing the logic of the assumptions dynamically,[10] there are two logically contradictory tendencies working against one another. On the one hand we are continually removing humans from our automated systems whenever we can. On the other hand, we are striving to formulate an ethics of our human relations to such systems. This contradiction renders our attempt at an ethics halfhearted at best, for if we succeed in removing ourselves, then why do we need an ethics?

These contradictory assumptions explain the ambiguity in meanings of the 'Human…' terms however. To the extent that the AI or automation can take over human roles, the human gets measured as a component of some complexity in the system. To the extent that AI or automation has not yet impinged upon more complex human capacities, the human is measured in terms of having degrees of oversight and control from a position 'outside' the system. Often the two are confused because automated systems have components and phases of varying complexity.

Further potential confusion on the 'oversight' side of the issue is mentioned in Mihály Héder's clear discussion on what constitutes genuinely AI-specific concerns for AI Ethics guidelines. On the one hand human oversight of the type categorized by the 'Human …' terms seems to be rightly applied to the advanced levels of autonomy which is specific to AI (69). On the other hand, our wish for AIs to have autonomous control, is quite directly and paradoxically opposed to our wish for control of the AI in terms of unwanted consequences which we nonetheless cannot foresee (63). Reconciling these two urges may well be impossible then, rendering terms such as 'Human …' which propagate an outlook of control vs. resignation in face of the advance of AI and related technologies, very doubtful practically in an ethical sense, although they may well be or seem useful in a technical sense.

An ethical consideration of the 'Human…' terms and a rethink regarding their use *at all*, is thus needed for several reasons. The confusing multiplication of the terms is unclear and unsuitable for ethical development. The attempt to generalize selected aspects of the intent behind the terms, such as oversight, is difficult to practically implement and bypasses the important ethical discussion of whether and how humans are 'parts' of autonomous systems. Finally, the originating uses of the terms have embedded within them a fundamental logical contradiction between two assumptions: the human, viewed as component, should be removed from automated systems vs. the human, viewed as overseer, should build up an ethical relation to such automated systems.

4. Toward New Terms

The impression that humans gradually lose ground against machines is not new. "For a while I was afraid that Apollo might be one of the last battlefields on which human race took up arms against the encroachment of machines." (Shea 7) For Shea it is better that humans give up tasks which automated system could do. His own catch phrase for the intent of the 'Human …' terms was: "man optimized in the system," which harkens back to Birmingham's general outlook, although perhaps it is a call for technology designed to promote higher creative human capabilities. If the latter, can that be achieved by viewing the human as 'in the system,' with its implication of an 'outside the system' state?

Reflecting further on the 'in' and 'out' characterization of what could be called a human *loss of purpose* suggests a possible solution. This loss of purpose is due to a forgetfulness about our roles in technology. Technological advances are extensions of ourselves, but in a much deeper sense than being mere tools used as means, and they have ethical implications accordingly.

[10] For more on using logic dynamically to engage ethics, see Marc M. Anderson, *Hyperthematics: The Logic of Value*. NY, SUNY, 2019, and appendix D in particular.

Some insights from Bruno Latour can help us get a better sense of the weakness of the component and oversight approach to human- 'machine' relations. Latour applies a viewpoint which resonates deeply with Whiteheadian process thinking, in order to unpack our unconsidered human tendency toward technological objectification. In *La Fin Des Moyens* he suggests that we remember how the objects of our technologies are really 'folded' abstractions whose true origins lie in the intermediary non-human natural processes and ancestral technological processes of current technologies (Latour, 43). Our technologies are folded layers of complexity built upon simpler and older technologies and upon simpler human approaches which created those technologies. They are built upon a vast number of 'materials' appropriated from their original processes, whether plants, or mineral.

Latour gives the example of a hammer sitting there unproblematically as a seemingly compact and complete technological object, a 'neutral' tool, ready for use. Such technological objects are never neutral however, as Latour notes. The hammer is linked to the minerals, the tree from which its handle came, the iron smelters, the factories, and so forth, which have made it possible. The contemporary AI's 'veins' are also filled with the hard-won 'electric blood' of a human manipulation of energy and stored data, and supported by the minerals in the microchip, the oil in the plastics, the factory which makes the chip, and the water used in the process.

Latour (52) suggests that the task of ethics is to remember these forgotten sources of technical objects. Ethics slows things down and helps us to stop treating technological intermediaries as mere means, but rather as potential entities of a communal world which we build together. In *What is the Style in Matters of Concern?* Latour has argued that this participatory creation of the world is involved in the creation of our objective facts as well, in the labors of scientific communities (2005). We do not need to enter into that debate. We do need to recognize that whatever the precise role of humans in the world's objectivity and whatever the role of technological and simple materials in our technologies, it is unhelpful ethically to view the technological human as stripped of human relations, including ancestral precursors of work organization and relations to the human's own life as a process with purpose. In a variation of the contradiction above, *a* human as component is a human being 'folded up' in ways which work against the ethical recognition of that human. The gradual removal of all such sufficiently 'componented' humans, works, in turn, against the ethical recognition of humanity in relation to technology.

Oversight is not a sufficient defense against this component viewpoint, being an unclear notion slated to disappear eventually as the human is removed. A better response is to begin by evaluating the roles of individuals with regard to technology development processes. From there, and ideally tailored to our evaluation for each individual, we can go on to expand these individuals out of a mere component role. This will shift us from an attitude which accepts oversight from *outside* of the technology into an active participation *in* the creation of the technology. That participation would lay the foundation for a true ethical engagement of the technology. Being a mere component would thus be transformed into being *included* actively as a member of a community of humans engaged in technology producing actions. Oversight would then be transformed into actively *guiding* ourselves as a community of humans engaged in technology producing actions. Transforming the negative perspective of removing humans when possible, would be transformed into *what humans bring by persistent participation in the human community of technology*. The scale present in Table 2 outlines some ranges of this new outlook that we could begin with.

Table 2. Ethically Oriented Scale for Human Involvement in Technology Producing Communities

Term	Human Participation in Technology Producing Communities
Inclusion	1. *Replaceable* – technology community views a human as replaceable by any other human
	2. *Experienced* – technology community views human as experienced in multiple activities

	3. *Unique* – technology community develops technology around a human's unique skills and life context
Guidance	1.*Tester* – human's use of system is merely 'registered' (passive)
	2.*Trainer* – human's suggestions regarding system are implemented
	3.*Designer* – human designs complex parts of the system with others
Persistence	1. *Brief* – human initiates or contributes to an abstracted phase of a technology community
	2. *Sustained* – human contributes to the actions of a technology community up to the completion of the technology community's goals
	3. *Evolving* – human contributes to the process of a technology community as it overlaps and weaves into different technology communities beyond itself

Our scale attempts to capture *the relation of the human to the human community developing the technology,* rather than to the technology alone as a distinct 'object.' Representing the categories by capital letters in the order IGP, the human relation in any situation can thus be indicated by a simple subscript recording its rating under each of the three categories, with the lowest rating being IGP_{111} and the highest being IGP_{333}. Further categories for more subtle aspects of the relation could be added, including the ethical recognition of the intermediaries behind and within our technologies, and the inclusive participation of those who have firsthand experience of such intermediaries, e.g. indigenous peoples.

Practically speaking, the scale would be used for each group of humans who have the same similar actions in the community's project, although it could be applied to individual humans as desired. Establishing the context would consist at minimum of asking: who are the humans involved and how can we group them? We can be flexible here and expand or focus to choose the scope of our evaluation, depending on our needs and resources. Sometimes this might mean limiting ourselves to a sub-community of some larger technology development community such as the mining community, or the managers of a development project.[11]

Example 1 – A Crowdsourced Supervised Machine Learning project

We might apply the scale to a project of *CrowdSourced Supervised Machine Learning*. Here the community might consist of requesters/developers, microworkers, and users. The micro-workers, might rate between IGP_{111} and IGP_{113}, i.e. for **inclusion** 1-*replaceable* – the development community views them as *replaceable by any other human* ('anyone' could do the microworker tasks), for **guidance** 1-*tester* – their use of the system is merely 'registered' passively (they are merely providing data), and finally, under **persistence**, their contributions might be 1-*brief* – only used in some small part of the technology development project in question, 2-*sustained* – used in all of it, or 3-*evolving* – used beyond it in future projects (in the form of the dataset).

The requesters/developers, meanwhile, might fulfill the highest levels of all three terms, so that their experience would rate as IGP_{333}, i.e. for **inclusion** 3-*unique* – the technology is developed around their unique skills and perhaps life contexts (they have trained toward and are specialized in certain areas), for **guidance** 3-*designer*

[11] The website https://anatomyof.ai and Kate Crawford's *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, are wonderful resources through which one can get a sense of the vast and intricate web of human communities which ultimately should be considered in technology development. That understanding in turn can help in choosing the scope of local or broader evaluations of human involvement such as we are suggesting with IGP.

– they design complex parts of the system with others (they design complex parts of the algorithm), and for **persistence** 3-*evolving* – they contribute to the process of the technology community as it overlaps and weaves into different technology communities beyond itself (their participation flows on into further projects which overlap with the Machine Learning project goal in question).

Finally, the end users, might be IGP_{122}, where such a user is for **inclusion** 1-*replaceable* – viewed by the technology community as replaceable by any other human, for **guidance** 2-*trainer* – the user's suggestions regarding the system are implemented (substantial user feedback is sought), and their **participation** is 2-*sustained* – contributing to the actions of the technology community up to the completion of the technology community's goals (they have an opportunity to provide the feedback suggestions throughout the lifetime of at least one version of the system).

Example 2 – Self Driving Car development

Application to a community developing *Self Driving Cars* can serve as a further example. The car owners who drive, the developers of the car's algorithms, the remote assistance drivers, and the self-driving pilots who monitor vehicle and surroundings can all be considered as part of the technology developing community.

The micro-workers, who are viewed as replaceable by the development community – 'anyone' could do these tasks – and who are merely providing data, but whose contribution in the form of the dataset might be re-used, might rate between IGP_{111} and IGP_{113}. The requesters/developers, meanwhile, might fulfill the highest levels of all three terms, bringing their unique abilities and contexts, designing complex aspects of the algorithm, and contributing to further diverse projects which overlap with the goal in question. Their experience would thus rate at IGP_{333}. Finally, the end users, might be IGP_{122}, where such a user is 'anyone,' but has an opportunity to provide feedback suggestions throughout the lifetime of at least one version of the system.

Application to a community developing *Self Driving Cars* can serve as a further example. The car owners who drive, the developers of the car's algorithms, the remote assistance drivers, and the self-driving pilots who monitor vehicle and surroundings can all be considered as part of the technology developing community. The car owners, being 'anyone,' and having input options, but not participating in a sustained way in the development of the system, might rate as IGP_{121}. The algorithm developers might rate between IGP_{222} and IGP_{233}, depending on what and how long they contribute to the project. The remote assistance drivers, who develop experience to 'advise' the self-driving algorithm in complex real-world situations, whose suggestions are implemented directly by the car, and whose contribution may or may not go beyond the project of improving the self-driving system, would rate at IGP_{222} or IGP_{223}. Finally, the self-driving pilots or 'sitters,' who have basic training but relatively passive response options – e.g. pressing a red button to stop the car – would rate at IGP_{112} or IGP_{113}.

It should be evident that the current 'Human …' notions of oversight are encompassed within the IGP scale but in a broader and potentially more detailed sense. Self-driving pilots as described above, for example, are in indeed 'in command' at IGP_{112} as HIC would categorize them – they can shut down the system – but from the point of view of human participation, rather than that of technical control systems, the ethical prospects of this oversight is poor indeed, as reflected in the rating, i.e. they are not 'really' in command as they would be at IGP_{333}, if they had participated in developing the algorithm, etc.

The emphasis is on bringing value to the experience of individual humans in their relation to technology producing communities by helping them establish a baseline of their current experience with those communities; i.e. the expansive process of reflection undertaken in deriving the rating is more important than the exact numbers arrived at. The rating then serves as an incentive for those communities to put the human

at the center.[12] From rating human participation in the community of technology, including that of AI, we might then go on to recast the notions of responsibility (persistence of human participation), privacy (limiting guidance to willing participations), diversity (encouraging expansive inclusion into the community of technology), and so forth.

5. Conclusion

We began with an initial exploration of the history of the 'Human …' terms and found them to represent a variety of sometimes confused conceptions of the human relation to automated systems. We argued that the addition of many news terms has made them unclear and the attempt at generalizing the terms as oversight is not practical and misses serious ethical issues. Most importantly, there is a logical contradiction inherent in the original intent behind the terms which works against efforts to develop an appropriate ethics. Drawing on Latour, we suggest that, just as the provenance of the technological object can be forgotten, the provenance of the human as component has also been forgotten in the 'Human …' terms. It can be recovered by expanding of the role of the human within the community creating the technology. We offer a new scale to grade this expansion.

In suggesting this new scale, we are not critiquing a practical use of the terms internally to particular technological environments, e.g. that of Machine Learning or Robotics development, although even here our considered impression is that current usages for such limited technical purposes are often inconsistent because the background of the terms is unknown to those using them. Rather we are attempting to draw the discussion out into the larger question of: what sort of scale might engage and cast light upon the wider human relation to technologies such as AI, so as to draw technological projects into a larger project of ethical community?

The prospects of this larger project of ethical community is intimately related to the various fields of human activity and the ideas, terms, and practices we use at ground level in those fields. Proximally, this line of research can be extended by closely questioning our many other technological 'catchphrases.' Where have these terms come from? What assumptions about human relations to other humans and to our technology do they bring with them? How do those origins color our often unreflective assessments of the contexts in which we use them? This line of research can also extend all the way to a more complex discussion on our future with technology. For example, what does the intent to remove humans from our technological systems eventually imply for the future of human work?

[12] The IGP does not directly address participation in malevolent technology developing communities, but assumes that participation as such is a precursor for ethical outcomes, and that such communities would be disclosed in the IGP rating that would capture their relation to the humans harmed by them in a sufficiently broad selection of the limits of the community.

References

Alper, Paul. "An Open-Loop Procedure for Process Parameter Estimation Using a Hybrid Computer." Theory of Self-Adaptive Control Systems: Proceedings of the Second IFAC Symposium on the Theory of Self-Adaptive Control Systems, September 14-17, 1965, National Physical Laboratory, Teddington, England. Edited by P.H. Hammond. Springer Science, NY. 1966. https://link.springer.com/book/10.1007/978-1-4899-6289-8

Anderson, Marc M. Hyperthematics: The Logic of Value. New York. SUNY Press. 2019.

Ayuso, Damaris M. "Topic Session on Discourse." Proceedings of the 5th conference on Message understanding (MUC5 '93), p. 345, 1993, doi.org/10.3115/1072017.1072051

Caid, W.R. and Webb Simmons. "Minicomputer Assisted Reprogramming System (Mars)" Rept. for 15 Nov 77-15 Nov 79, Defence Nuclear Agency, 1979.

Crawford, Kate. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale. Yale UP. 2021.

Brown, Gordon S. and Donald P. Campbell. Principles of Servomechanisms: Dynamics And Synthesis Of Closed-Loop Control Systems. John Wiley & Sons, NY, 1948. https://archive.org/details/in.ernet.dli.2015.166797/page/n5/mode/2up?q=d

Bennet, Corwin, A. "Some experimentation on the tie-in of the human operator to the control loop of an airborne navigational digital computer system." IRE-ACM-AIEE '57 (Eastern), 1957, pp. 68-71, doi/10.1145/1457720.1457732.

Batali, John. "How Much AI Does a Cognitive Science Major Need to Know?" SIGART Bulletin, vol. 6, 1995, pp. 16-19, doi/10.1145/201977.201985.

Birmingham, H.P. and F.V. Taylor. "A Design Philosophy for Man-Machine Control Systems." Proceedings of the IRE, vol. 42, no. 12, 1954, pp. 1748-1758, doi: 10.1109/JRPROC.1954.274775.

Cummings, M.L., et al. "The Impact of Human–Automation Collaboration in Decentralized Multiple Unmanned Vehicle Control." Proceedings of the IEEE, vol. 100, no. 3, pp. 660-671, 2012, doi: 10.1109/JPROC.2011.2174104.

Day, Dwayne A. "Invitation to Struggle: The History of Civilian-Military Relations in Space." Exploring the Unknown: Sellected Documents in the History of the U.S. Civil Space Program, vol. 2. External Relationships. Washington DC. 1996.

De Lemos, Rogério. "Human in the loop: what is the point of no return?" Proceedings of the IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '20), pp. 165–166, 2020, doi.org/10.1145/3387939.3391597.

Docherty, B. et al. "Losing Humanity: The Case Against Killer Robots." Communications of The ACM, vol. 42, 2012.

https://www3.nd.edu/~dhoward1/Losing%20Humanity-The%20Case%20against%20Killer%20Robots-Human%20Rights%20Watch.pdf

Fanni, R. et al. "Active Human Agency in Artificial Intelligence Mediation." Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good, pp. 84–89, 2020, doi.org/10.1145/3411170.3411226.

Gschwind, Robert T. "Control Dynamics of Human Tracking with a Viscously Damped Tracking Aid" Memorandum Report No. 2709, USA Ballistic Research Laboratories, pp. 1-28, 1976. https://apps.dtic.mil/sti/pdfs/ADA035455.pdf

Hammes, T.X. "Reality in Autonomous Systems: It starts the Loop" The Cove, 2020. https://cove.army.gov.au/article/reality-autonomous-systems-it-starts-the-loop (Accessed August 2021)

Héder, Mihály. "A criticism of AI ethics guidelines." Információs Társadalom, vol. 20, no. 4, pp. 57-73, 2020, doi.org/10.22503/inftars.XX.2020.4.5.

High, Peter. "Carnegie Mellon Dean of Computer Science On the Future of AI" Forbes, 2020. https://www.forbes.com/sites/peterhigh/2017/10/30/carnegie-mellon-dean-of-computer-science-on-the-future-of-ai/ (Accessed September 2021)

HLEG. Ethics Guidelines for Trustworthy AI. European Commission, 2019. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

Hopkinson, William C. and José A. Sepúlveda. "Real time validation of man-in-the-loop simulations." Proceedings of the 27th conference on Winter simulation (WSC '95), pp. 1250–1256, 1995, doi.org/10.1145/224401.224804.

Latour, Bruno. "La fin des moyens." Reseaux, vol. 18, no. 100, pp. 39-58, 2000. https://www.persee.fr/doc/reso_0751-7971_2000_num_18_100_2211

--- What is the Style in Matters of Concern?. Van Gorcum. Assen, Belgium. 2008

Loy, Patrick. "The method won't save you: (but it can help)." SIGSOFT Softw. Eng. Notes, vol. 18, no. 1, pp. 30-34, 1993, doi.org/10.1145/157397.157398.

Merat, N., Seppelt, B., Louw, T. et al. "The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications." Cognition, Technology & Work. vol. 21, pp. 87–98, 2019. doi.org/10.1007/s10111-018-0525-8

McLaughlin, Margaret L. et al. "The haptic museum." Proceedings of the EVA 2000 conference on electronic imaging and the visual arts. 2000. https://infolab.usc.edu/DocsDemos/eva2000.pdf

McRuer, Duane T. and Ezra S. Krendel. "The human operator as a servo system element." Journal of the Franklin Institute, vol. 267, no. 5, pp. 381-403, 1959, doi.org/10.1016/0016-0032(59)90091-2.

Naval Research Laboratory. "Report of Naval Progress." 1958. https://www.google.fr/books/edition/Report_of_NRL_Progress/z7I1AAAAMAAJ?hl=en&gbpv=1&dq=%22human+in+the+loop%22&pg=RA2-PA20&printsec=frontcover

Porathe, Thomas and Johannes Prison. "Design of human-map system interaction." CHI '08 Extended Abstracts on Human Factors in Computing Systems. Association for Computing Machinery, pp. 2859–2864, 2008, doi.org/10.1145/1358628.1358774.

Reynolds, Paul F. 1988. "Heterogenous distributed simulation." Proceedings of the 20th conference on Winter simulation (WSC '88), pp. 206–209, 1988, doi.org/10.1145/318123.318190.

Rissland, Edwina L. and Jody J. Daniels. "A hybrid CBR-IR approach to legal information retrieval." Proceedings of the 5th international conference on Artificial intelligence and law (ICAIL '95), pp. 52–61, 1995, doi.org/10.1145/222092.222125.

Sanders, Daniel S. "Social Work Concerns Related to Peace and People Oriented Development in the International Context." The Journal of Sociology & Social Welfare, vol. 15, no. 2, pp. 57-72, 1988.

Shea, J. "Systems Engineering for Manned Space Flight" 2nd Manned Space Flight Meeting. American Institute of Aeronautics and Astronautics. NY. 1963.

Stieber, Michael E. et al. "Control of Robotic Systems on the Space Station." IFAC Proceedings Volumes, vol. 31, no. 33, pp. 89-94, 1998, doi.org/10.1016/S1474-6670(17)38392-1.

Stotz, Robert. "Man-machine console facilities for computer-aided design." Proceedings of the May 21-23, 1963, spring joint computer conference (AFIPS '63 (Spring)), pp. 323–328, 1963, doi.org/10.1145/1461551.1461590.

Stouch, Daniel, et al. "Coevolving collection plans for UAS constellations." Proceedings of the 13th annual conference on Genetic and evolutionary computation (GECCO '11), pp. 1691-1698, 2011, doi.org/10.1145/2001576.2001804.

Wagner, Markus, "Taking Humans Out of the Loop: Implications for International Humanitarian Law." Journal of Law Information and Science, vol. 21, 2011. https://ssrn.com/abstract=1874039

Wixon, Dennis and John Whiteside. 1985. "Engineering for usability (panel session): lessons from the user derived interface." SIGCHI Bull., vol. 16, no. 4, pp. 144–147, 1985, doi.org/10.1145/1165385.317484.