AI and Society manuscript No.
(will be inserted by the editor)

The Model Gap

Cognitive Systems in Security Applications and their Ethical Implications

Tobias Matzner

Received: date / Accepted: date

Abstract The use of cognitive systems like pattern recognition or video tracking technology in security applications is becoming ever more common. The paper considers cases in which the cognitive systems are meant to assist human tasks by providing information, but the final decision is left to the human. All these systems and their various applications have a common feature: an intrinsic difference in how a situation or an event is assessed by a human being and a cognitive system. This difference, which here is named "the model gap", is analyzed pertaining to its epistemic role and its ethical consequences. The main results are: 1) The model gap is not a problem, which might be solved by future research, but the central feature of cognitive systems. 2) The model gap appears on two levels: the aspects of the world which are evaluated and the way they are processed. This leads to changes in central concepts. While differences on the first level often are the very reason for the deployment of cognitive systems, the latter is hard to notice and often goes unreflected. 3) Such a missing reflection is ethically problematic because the human is meant to give the final judgment. It is particularly problematic in security applications where it might lead to a conflation of descriptive and normative concepts. 4) The idea of the human operator having the last word is based on an assumption of independent judgment. This assumption is flawed for two reasons: The cognitive system and the human operators form a "hybrid system" the components of which cannot be assessed independently; and additional modes of judgment might pose new ethical problems.

Keywords model gap · cognitive systems · security ethics · smart CCTV · drone · UAV

Tobias Matzner
Wilhelmstraße 19, 72070 Tübingen, Germany
Tel.: +49-7072-2977988
Fax: +49-7072-295255
E-mail: tobias.matzner@uni-tuebingen.de

1 The model gap

Regarding security, we constantly evaluate our situation according to highly context-dependent concepts. This pertains to two levels: social and personal. On the social level various contexts like airport, train station, market, private accommodations, etc. can be distinguished regarding both the prevalent security expectations themselves and the value security has compared to other values like freedom, privacy, or justice. On the personal level, in each of these contexts, everyone has their own prospect of security. For example, a rather dirty street lined with graffiti makes some people feel insecure, yet it can be the sought-after neighborhood for others. In a similar manner, everybody places security at a different position regarding competing values. These context-dependent factors contribute to a variety of differing perceptions on what counts as threat to security in a given context.

Of course, the social and the personal levels are related. The exact nature of this relation is debated in various scientific discourses, which I do not want to get into here. My remarks are just intended to highlight that it does not suffice to reflect the prevailing views or social standards or to presuppose a rather homogeneous and settled view in most contexts.

Smart[1] security systems introduce one or more additional evaluations of the context. For the purpose of this paper, these systems in are conceived of in a rather broad manner: they use algorithms that evaluate sensor data (camera images, sound recordings, movement sensors, etc.) with the aim to detect events that are noteworthy for the operators. The analysis is focused on systems that detect particular

[1] I follow the common use of the word "smart" as in "smart security system" or "smart CCTV" both in public and scientific discourse. Yet, it is important to mention that this choice of words can contribute to the very misjudgment of security technology that is discussed in this paper.

events or situations, but that leave the decision whether to act and what to do to human operators. In what follows such systems will be called "assisting cognitive systems". Such a combination can be found in many security applications that are currently being developed and deployed in ever increasing numbers. Some intensively discussed examples are smart CCTV, full body scanners which automatically detect suspicious areas (hiding the naked image), biometric identification like face recognition, or command and control for unmanned aerial vehicles.

Having human personnel evaluate the output of cognitive systems before deciding on further actions to be taken is often considered to solve many ethical problems of automation. (See for example Macnish (2012) on the case of smart CCTV.) This paper argues that such a combination of human operators and cognitive systems *per se* has some ethically problematic features—apart from the many problems that particular uses of such a system might pose. Those problems rise from the different evaluation of a context by human beings and cognitive systems.

Cognitive systems need a codification of that which they are meant to detect. This can be done in a positive fashion, determining certain events and triggering a signal in every case in which some or all criteria of this description are met. The other approach is negative in the sense that the system codifies normal activity for the given contexts and triggers the signal at every deviance from this normality.

Either of these approaches applies a *model* of the context. It is used to represent the activity in the context and to distinguish the activity to detect. In terms of computer science, what I call model consists of certain "features"[2], i.e. a selection of particular features of the sensor input that are evaluated, and a method for pattern recognition to decide which combinations of these features are considered to correspond with the event to detect (Bishop 2006, 1). Such a model could be based on movement trajectories of objects, three dimensional simplified models of human bodies, descriptions of the context in a formal language, exemplary sensor data, and many more. Usually, a smart system would combine several of these approaches, often by layering them. Thus, for a higher layer the "features" could be the hypotheses of lower layers, for example detected objects or persons, or predicted movement trajectories. To simplify things, below "the model" refers to the level on which the decision whether an alarm has to be triggered takes place.

The functioning of the smart system can be described in two steps: As a first step a representation of the context in the respective model is generated based on the sensor data. For example, several pixels in a camera image are interpreted as object; then several such objects in consecutive images are interpreted as the same and united into a movement trajectory; these data then could be classified by an algorithm as, for example pedestrian, cyclist, and car. (Usually, either of these stages would consist of several further intermediate steps.) In a second step this model allows for searching particular events based on their representation in the model. For example the system could detect that a car hit a cyclist and trigger an alert.

Research and development of smart security technology has yielded an enormous variety of models used in this area. Nearly all known approaches in machine learning and pattern recognition have been proposed for security applications. (See Hu et al (2004) for an overview.) Here, I do not want to go into the details of these approaches. (However, as I argue below, an ethical assessment of a particular system needs to take these specificities into account.) Instead, I want to focus on a common feature of all these approaches: the models that are used differ from the perception of the respective situation by the operating personnel. They differ as well from the perspective of those persons that are confronted by the consequences of the use of the system, e.g. people passing a body scanner or entering an area under surveillance by smart CCTV.

This difference does not only pertain to the social and context dependent factors that influence what is considered as a security threat, which I mentioned in the introduction. There is also a difference in the very terms based on which the situation is perceived. For example, Irma van der Ploeg describes how the concept of "identity" changes with the introduction of biometric technology (Van der Ploeg 1999). And even (supposedly) basic and material things like a human body change their meaning when processed by biometric cognitive systems (Van der Ploeg 2005). Van der Ploeg here points to a decisive feature that also applies to cognitive systems: To provide information pertaining to a human concept like "identity," different aspects of the world are evaluated. Using the EURODAC[3] database as an example, she shows how identity based on papers and the authority of the issuing offices transforms into identity based on bodily features like fingerprints. "Identity" should not be seen as something fixed, which then is determined by various means, but a different kind of identity is established by the introduction of biometric technology (Van der Ploeg 1999, 300). Of course, using a system that evaluates other aspects of the world was a conscious decision at the introduction of EURODAC and biometrics to counter some perceived shortcomings of document-based identification. Yet, the accompanying change of identity itself is not as easily

[2] In this context, the first processing steps of a cognitive system are often called "feature extraction."(Bishop 2006, 2) Yet, I want to stress that this is already an interpretation *creating* a new description, instead of *extracting* something that is already there.

[3] The EURODAC is a central database using fingerprints to identify all asylum seekers and illegal migrants in the European Union.

reflected – and even less the changes of underlying concepts like the body.

This is the first important point comparing the evaluation of a situation by human beings and cognitive systems: a concept is evaluated concerning different aspects of the world and thus the concepts themselves change.

But there is a second difference: Cognitive systems do not only evaluate different aspects of the world, they also process this information in a completely different manner. For example, applications of biometric identity started long before digital information technology with human beings comparing fingerprints or faces with photographs. Still, the introduction of fast computing and large databases led to changes not only in quantity but quality, as those described by Van der Ploeg. But even if it was feasible to have human beings doing all the search work for fingerprints in databases, they still would by doing something different than cognitive systems do: Simply because a fingerprint or the photograph of a face, or a person showing aggressive behavior modelled in a cognitive system is something else as it is for a human being. As sketched above, a cognitive system deals with certain features based on sensor input and the things or events to recognize correspond to certain sets of those features.

To illustrate the difference: the infamous "eigenface" approach (Turk and Pentland 1991), one of the first promising face recognition algorithms, treats digital photographs of faces as vectors, i.e. a list of the greyscale intensity of each pixel. By a mathematical evaluation some decisive components of a large training set of such "faces" are determined: the "eigenfaces". (In mathematically correct terms: The eigenvectors are determined using principal component analysis.) The faces to be recognized are stored as a combination of such eigenfaces. In very general terms, some patterns are determined, which can be found in many of the photographs. Then the images are approximated by combining these patterns. So instead of storing the entire set of photographs, each face is represented just by a combination or overlay of these patterns. So all that needs to be stored is a few values determining the patterns or eigenfaces and a "weight," i.e. the amount each eigenface contributes to the image. Since the number of eigenfaces is significantly smaller than the number of pixels (the original approach of Turk and Pentland (1991) used seven components to represent images with 65365 pixels) this method allowed the efficient storage and processing needed for recognition.

So, in this case, a "face" in the cognitive system is just a set of seven values that get their meaning pertaining to a sophisticated mathematical evaluation (principal component analysis) of a set of training images, which are again treated as a mathematical entity (a vector space).

Those conceptual differences might be seen as a shortcoming of "smart" systems, considering marketing speech, research proposals, and political discussions lauding the detection abilities of the respective systems in terms of human descriptions or capabilities. Yet, it is important to note, that this is not the case. If cognitive systems promise to be of great help to human operators, it is *particularly because they evaluate other aspects* of the context and *process this information differently.* Thus they provide data a human being could not have obtained (e.g. object detection in the invisible spectrum) or just could obtain with a lot of time and effort (e.g. identifying a person in various video recordings). I introduce the notion *model gap* for this difference in the evaluation of a situation by human beings and a cognitive system.

Let me repeat, this model gap is intrinsic to the expected functionality and benefits of using cognitive systems as assistance to human operators. Thus, when we require that smart systems evaluate a situation as similarly as possible to the perception of a human being, we do not require that the systems learn the same concepts or features of the context. We want that the systems trigger an alarm only when that human being would trigger the alarm, but *based on a different evaluation* of the situation. This is something else as saying the system should only trigger an alarm when the event it is meant to detect really happens. As I remarked above, the question whether whatever is happening amounts to a security threat is highly dependent on the context and the individual asking it. Thus, a cognitive system can only be made to approach the perception of a particular observer in a given context. So reality is no neutral ground on which to functionally eliminate the conceptual difference of cognitive system and human beings.

This conception differs from the "social-technical gap" as prominently defined by Ackerman (2000, 180) as "the great divide between what we know we must support socially and what we can support technically." While Ackerman talks about a lack of technical systems as compared to the human social world, the model gap is about a difference that is the very reason why assisting cognitive systems are deployed. It does not diminish its functionality but enables it in the first place. This means that the potential problems posed by the model gap are intrinsic to assisting cognitive systems.

2 The problems posed by the model gap

From an ethical perspective, the model gap potentially yields several problems. The descriptions of what a smart system does commonly refer to human concepts: The system is meant to detect "suspicious persons", "abandoned objects", or to find "known criminals" in a face database. It signals "aggressive behavior", "violations of access rights", or "abnor-

mal events". Regarding assisting cognitive systems, it is of great importance that the operators know the difference between such descriptions of the situation and the model used in the system: it is the human operator that is meant to judge the output of the system and thus should know what this output means—and what it doesn't mean. This pertains to the difference established by Van der Ploeg: While a cognitive system is deployed because it can evaluate other aspects of the world, the changes of the related concepts are easy to miss. The same applies to the difference in processing and thus the reflection of what a signal from the system entails.

Reflecting the model gap is particularly important in security applications—and to an extent regarding safety as well—because concepts in these contexts are normatively charged. They describe negative events, acts that should be averted, persons that should be excluded—or even killed. Thus, the danger arises to interpret the signals of the system as pertaining to illegal, immoral, dangerous, or otherwise negative situations—when in reality the signal just means that the model has reached a state which is meant to represent such a situation.

In this way, the problems of such a conflation play out on two levels. On the first level there are different *descriptive* conceptions of what is going on: those by human beings and those by cognitive systems. Already on this level, missing the model gap can lead to ethically problematic actions by the personnel. Furthermore, when the system is used to provide assistance to human operators, the signals might even mean less than the event to detect has happened (as far as the cognitive system can discern it): Knowing that there is an operator to check, the system might be designed to issue a signal already when some indicators are matched or the probability of a noteworthy event has reached a certain threshold. Consequently the signal means that the situation should be checked *if* something requiring action by the operators is happening—not that it *is* happening. The second level comes into play since the event to detect is something that has a normative value. A missing understanding of the model gap may lead to a conflation of something being not normal *descriptively*—based on statistics or pre-given rules—with something that is *normatively* wrong. This is the case particulary for learning systems, that build a statistical model of the descriptively normal sensor input and trigger an alarm at significant deviations—since not everything that is exceptional is problematic.

This problem is exacerbated by the various perspectives on security that exist in every context, as explained above in the introduction. If the model gap is not consciously reflected, an operator might liken the alarm to her or his conception of the situation: if the system singles out a person as "suspicious" this will mean whatever "suspicious" means for the operator—yet now enhanced with the supposed objectivity of the machine. Other perceptions of the same situations thus might be ignored or devaluated.

In a similar manner, people confronted with the consequences of an alarm will reflect this based on their own perception. If the system is perceived to counter "threats to security" (as compared to signal certain probabilities for threats) many alarms and further actions (e.g. being searched after passing a body scanner) will be considered to be wrong. Consequently, the system will be perceived as malfunctioning, useless, discriminatory, or threatening, if people do not know that the alarm does not actually mean the existence of a threat. It just means that an algorithm detected an event that *might* be a threat and thus should be further analyzed by security personnel. Of course this effect is increased if the personnel don't reflect the model gap either.

The problem just discussed, that cognitive systems are based on a descriptive model of situations or events that do not coincide with their normative evaluation, does not mean that such a system is normatively neutral, in the sense that the system "just" describes the context and the human operators judge it. As Bowker and Star (2000, 135) note, "values, opinions and rhetoric are frozen into code". A similar point is made by Brey (2000) and Introna (2005). Nearly all levels of system design allow for the influence of normative presuppositions: The choice of features and the model decides what is "visible" to the system. For example, a system built to recognize abnormal behavior in CCTV images based on motion might not be able to distinguish a person stumbling and getting hurt from a homeless person sitting down—in either case the movement stops where usually persons don't linger for too long. A further problem is the choice of training data for statistical models. These data represent either the normal case or the events to detect. Here, the concepts and values of the persons responsible for the choice clearly influence the system. These presuppositions, however, are hidden behind the functionality of a technical system. Thus, they are no longer presuppositions in human terms and can only be discerned by either a minute assessment of the system or by thorough knowledge and reflection of the design and development processes.

This problem is complicated since not in all cases training examples and other data provided necessarily comply with the model. Bowker and Star (2000, 156) describe how the perception of a coding as inappropriate or ill-fitting can lead to mechanisms of circumvention or evasion like entering wrong or bogus data since no fitting representation can be found. As I argued above, such a perception of the system might likely result from a missing reflection of the model gap.

3 Reflecting the gap

As lined out above, the existence of the model gap is at the very core of assisting cognitive systems. Thus, to interpret the signals of the system correctly, the operators have to understand the model gap. This means that the model has to be translated into human terms. So for example, the signal of a "suspicious behavior" might be translated as: the movement trajectory of this object deviates by a certain threshold from the average movement of objects in this scene regarding a particular mathematical measure that expresses the "distance" between movement trajectories. Respectively, the "object" in case is the hypothesis of a particular pattern recognition algorithm that this group of pixels represent a certain object.

As I noted above, state-of-the-art systems in research and on the market usually make use of many complex components in several layers or steps. For various reasons, not all of this can be known in detail by operators: First of all, the required prior knowledge of engineering and computer science cannot be presupposed. Furthermore, a complete disclosure of the functionality of the system makes it easier to circumvent and thus threatens its functionality. And finally, the details of the models and algorithms usually are the unique features of assisting cognitive systems and consequently expensive trade secrets, which would not be made widely available.

Apart from these concerns, which prohibit the disclosure of details about a cognitive system to the operators, some approaches in pattern recognition do not allow for an easy translation of their working in human terms. It has been noted as a classical mistake of early AI research to assume that human intelligence or the human brain works like a very sophisticated symbolic computer—where the reduction of intelligence to the brain already is part of this mistake (Dreyfus 1992). Later insights led AI research to correct these mistakes, going for complex statistical models or subsymbolic approaches like neural networks, fuzzy logic, or genetic algorithms – often subsumed under the label "computational intelligence" (Bezdec 1994). Those approaches do not try to use human concepts but to create an adaptive behavior that can achieve human defined goals (Eberhart 2007). I have already pointed to the problem of whose evaluation of a situation or an event should be seen as reference for an appropriate functionality. Subsymbolic and many statistical systems, however, bring with them the additional problem, that they can only be evaluated by their output but it is impossible to translate the meaning of single parameters into human terms (Robinson 1992; Stamou et al 1999). In other words: it is hard to impossible to understand how a specific output was created.[4]

Despite these problems, a certain understanding of the model gap by the operator is needed, since it creates the problems summarized in section 2. For the various reasons just mentioned, however, this cannot be based on a detailed knowledge of the inner workings of assisting cognitive systems. But there are other reasons against such an approach that would require security personnel becoming IT experts: In an article on human-computer interaction Brey (2005) distinguishes the perspective of a programmer, who deals with computers on the algorithmic level from the perspective of the user who works with the computer on the functional level. For the programmer, a computer is and always will be an information processing tool—because essentially even our sophisticated IT infrastructure consists of nothing else but fast calculators. Yet, for the users, computers offer a wide range of functions that need not be related to information processing and calculation: making music, painting images, playing games (293-4). So what a computer does is not a matter of its inner workings for the user. What counts is "the purpose assigned to them by designers and users" (394).

It is on this functional level, that the reflection of the model gap has to play out. The assisting cognitive system has to change from a system that for example detects criminal behavior to a system that signals increased probabilities for criminal behavior. It should become a system that does not "look at the world" but which processes only certain features of the sensor input and processes them in a way that has implications in terms of what is "visible" or "invisible" to such a system. So for example it should be known that the system analyzes movement in space, gestures based on the spatial relation of skin-colored areas, or objects as described by a certain list of concepts; and that the alarms are triggered using deviance from statistical average, rules in a formal language or thresholds of likelihood based on a predictive model.

In general, it should be known what the functional implications of the model gap are, in which sense the assisting cognitive system changes the relevant concepts – or pertains to different concepts. This, however, is not only a matter of conscientious reflection and training of the operators. Reeves and Nass (1996) have shown, that the way computers and other media are introduced and represented by their description, hardware design, user interfaces, etc. have an influence of how they are seen by the users. In an experiment which is relevant for assisting cognitive systems, computers have been established as "team mate" as compared to technical tools. As a result, the human members of the

[4] To include those systems into my view, I use the word "model" rather comprehensively, in the sense that not every part of a model must have its counterpart in the world.

team considered the computer to work more like themselves. They "thought the computer solved problems in a style more similar to their own." (Reeves and Nass 1996, 156) Furthermore, they found that there was more agreement between the computer and themselves, information by the computer was considered to be more helpful. This even led the human team members to change correct results to be closer to those of the computer. While Reeves and Nass draw some promising perspectives of better human-computer interaction from these results, they also results show how easily human perspectives and computational results are conflated.

Thus, ideally, a reflection of the model gap on the functional level should be part of the ethically responsible development, marketing, installation, and support of smart security systems. Based on such efforts, a conscientious operator can estimate the difference between the signal of the model and the events in the context. This difference makes additional checks by the operator necessary to decide on further action: if we do not want completely new concepts being applied to a situation, we need the human judgment to make the events detected by assisting cognitive systems mean, what they are intended to mean: for only human judgments deals in these concepts.

The necessity of such additional human perception of the situation before measures are taken of course is the rationale of designing smart systems that are meant *to assist* human operators. Yet, it is important that this rationale and all its aforementioned implications be known by the operators. With a grain of salt, this can be summarized as such: Assisting cognitive systems should not be seen as operating on the same level as human beings—their anthropomorphizing descriptions notwithstanding. They should rather be considered as sophisticated sensors: Like a camera or a microphone they provide data that *must be judged* by the human operator. They do not provide a judgment in human terms.

4 The problems of independent judgment

In the aforementioned article Brey (2005) makes a similar distinction: building on ideas of Norman (1993) he describes computer systems as "cognitive artifacts" that can either "replace", "supplement", or "enhance" human capabilities. The results of the last paragraph clearly indicate that assisting cognitive systems should be seen as a cognitive enhancement and not a replacement of some human capabilities—or even human beings in general. That reflective insight alone, however, does not solve the problems. Brey refers to the concept of a "coupled cognitive system" developed by Clark and Chalmers (1998) and Clark (2001). He describes this as "the linking of a human being with an external entity in a two-way interaction that includes information input from this entity and epistemic actions towards it"(Brey 2005, 389). While the notion of "interaction" already indicates a certain entanglement of human operator and cognitive artifact, Brey goes on to argue that some of those systems are best described as "hybrid cognitive systems." In those systems both parts human and machine are considered to be only "semi-autonomous information-processing systems"(Brey 2005, 392). Other authors, most notably maybe Haraway (1991) using the "cyborg" trope, have argued against even distinguishing human and non-human parts in such a hybrid system. Both views entail that a cognitive system assisting human personnel *cannot* mean that the operators get some "proposals" from the system which they then evaluate or assess *as if there were no such system.* In others words, it cannot mean that the human operators check whether the system "got it right."

Above I argued that without assessment by the operators, no final judgment what happened can be had. And even less can a normative judgment be had based only on the signals of a smart system. To this end, I lined out, a reflection of the model gap is necessary. This reflection, however, is not to establish the *independence* of the human operator. Such a reflection must not be seen as protecting the human operators from "misunderstanding" the system and thus independent judgment being derogated. On the contrary reflecting the model gap means to recognize the *dependence* or entanglement of operator and assisting cognitive system. It means that one cannot simply use one's own perception or one's own concepts of the situation to assess the output of the system. Consequently, my argument in the last section on "reflecting the gap" is not meant to drag the human out of the "hybrid system." It is meant to foster the recognition of being part of such a system – and that this leads to conceptual shifts in the assessment of a situation or an event. Taking up the result of the last section, that an assisting cognitive system is more like a sensor, the output of such a "sensor" can now be described as an intermediate result or internal signal of a hybrid system. Thus, the reflection of the model gap means to recognize the implications of a cognitive system being part of such a hybrid system. In comparison, getting the model gap wrong would mean to consider human and cognitive systems as two parallel assessments of the situation, where one could correct the other—in the cases considered so far: the human correcting the machine.

This has strong implications for the idea of the human operator having the final decision—and being responsible. This idea is motivated by the notion of a human being that can detect mistakes made by the cognitive system and only react if this is "really" necessary. Such a detection of errors would make it necessary to be able to assess the situation

as if there were no cognitive system and then compare the results. This possibility, however, is by no means a given.

A particularly aggravating instance of this problem is faced by operators of unmanned aerial vehicles (UAVs) in combat situations. It is the human operator that fires the arms, but the only resources to judge the situation are the sensor data provided by the UAV. Furthermore, these data would be of no use to the operator without automated processing due to both their amount and quality (Parasuraman et al 2007). Since research has shown that reliable automation can reduce pilots' workloads substantially (Dixon et al 2005; Dixon and Wickens 2006), further development is likely to increase the dependence on automated systems—which makes operation in case of failure of these systems impossible. Particularly the signal delay between the control room and the UAV makes reliance on smart algorithms an indispensable part of such systems (de Vries 2005). To put the problem pointedly, the danger arises that the operator will fire when the smart systems of the UAV prompt to do so.

In civilian security contexts, for example video surveillance, the dependence of the operators is not as strong and their actions often (but not always) have less serious consequences. In case of an alarm, smart systems can make the input data like video or audio recordings available to the operators. Some systems even provide additional sensors for the operators to reevaluate a situation that lead to an alarm.[5] And in many contexts, for example subway and train stations, airports, public places, etc., the situation can be personally checked. Yet, such an appearance of security personnel, which is only meant to check the signals of the system can be perceived quite differently: On the one hand, it might lead to the impression that something worrying is happening, or that one is in danger. On the other hand, it might be perceived as annoying, or even discriminatory. The latter is particularly the case if this happens regularly, for example a person with a prothesis that triggers the pattern recognition system of a body scanner.

Yet, already on the technical level (i.e. without security personnel doing personal checks) each additional assessment means a deeper intrusion into the privacy of persons by making existing sensor data available to human operators or by gathering, providing, and potentially recording additional data. In some cases, privacy by design means particularly that the data that would be needed for further checking must not be disclosed. Such a case are body scanners at airports, where the image of the naked body cannot be seen, and only suspicious areas are highlighted on schematic representation of a human body. Thus, even where additional modes of judgment are possible, they might have ethical worrying consequences.

This means that smart systems do not only face the trade-off between good, efficient functionality and ethical problems as it is often perceived. There is a trade-off between two ethical requirements as well: a reflected and careful interpretation and evaluation of the signals of a smart system and accordingly the prevention of ill-founded judgments on the part of the operators on the one hand, and the protection of privacy and the right to autonomous, unimpeded acting on the other hand. This is a rather delicate issue because the ill-founded judgments usually lead to similar or worse intrusions than the re-evaluation of the alarms. This means that for every system meant to assist human operators, the question what happens in case of an alarm must be confronted from an ethical perspective. Refraining to the position that the system "only" hints at potential threats is not enough. What such a "hint" means and which consequences it can have must be taken into account. Particularly, if the "hint" cannot be checked without taking ethically problematic action, the design of the system is questionable.

To sum up, the model gap is intrinsic to the application of cognitive systems as support for human operators. If this difference is not reflected, it can create ethically problematic judgments. This applies in particular to the operators of such systems, but also for the people that are affected by the consequences of the assessments based on such a system (for example people under surveillance). Concerning the last group, the presentation or description of a system and its purpose is highly relevant. Regarding the operators, a reflection of the gap must not be misunderstood as their assessment correcting or verifying the output of a cognitive system. To the contrary it means the recognition of their immersion in what can be called a "hybrid system" in Brey's terms. In particular this entails recognizing the difficulties of independent judgment. If additional modes of judgment are provided, this requires additional sensors or personal checks. Both of which, however, might pose ethical problems as well. Thus, the ethical requirement of reflected and careful judgment might only be achieved by means which are as well ethically problematic.

[5] Of course, it would be a mistake to consider video images or audio as unprocessed just because they are "less" processed than data from cognitive systems. Yet, having additional data that did not influence the outcome of a cognitive system can be an advantage. And in addition, our social skills and practices better reflect the epistemic implications of image and audio recordings, which are available for roughly a century now. That, however, does not mean that their application cannot cause all kinds of problems—as the vast literature from media and surveillance studies testifies.

References

Ackerman MS (2000) The intellectual challenge of cscw: the gap between social requirements and technical feasibility. Human-Computer Interaction 15(2):179–203

Bezdec J (1994) What is computational intelligence? In: Zurada JM, Marks RJ, Robinson CJ (eds) Computational Intelligence Imitating Life, IEEE Press, New York

Bishop C (2006) Pattern Recognition and Machine Learning. Information Science and Statistics, Springer

Bowker GC, Star SL (2000) Sorting Things Out – Classification and its Consequences. MIT, Cambridge, MA

Brey P (2000) Disclosive computer ethics. ACM SIGCAS Computers and Society 30(4):10–16

Brey P (2005) The epistemology and ontology of human-computer interaction. Minds Mach 15(3-4):383–398

Clark A (2001) Reasons, robots and the extended mind. Mind & Language 16(2):121–145

Clark A, Chalmers DJ (1998) The extended mind. Analysis 58(1):7–19

de Vries SC (2005) Uavs and control delays. URL http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA454251

Dixon SR, Wickens CD (2006) Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 48(3):474–486, http://hfs.sagepub.com/content/48/3/474.full.pdf+html

Dixon SR, Wickens CD, Chang D (2005) Mission control of multiple unmanned aerial vehicles: A workload analysis. Human Factors: The Journal of the Human Factors and Ergonomics Society 47(3):479–487

Dreyfus H (1992) What Computer Still Can't Do: A Critique of Artificial Reason. Mit Press, Cambridge, MA

Eberhart RC (2007) Computational Intelligence: Concepts to Implementations. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA

Haraway D (1991) A cyborg manifesto – science, technology, and socialist-feminism in the late twentieth century. In: Simians, Cyborgs and Women: The Reinvention of Nature, Routledge, New York

Hu W, Tan T, Wang L, Maybank S (2004) A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 34(3):334 –352

Introna L (2005) Disclosive ethics and information technology: disclosing facial recognition systems. Ethics and Information Technology 7(2):75–86

Macnish K (2012) Unblinking eyes: The ethics of automating surveillance. Ethics and Information Technology 14(2):151–167

Norman DA (1993) Things that make us smart: defending human attributes in the age of the machine. Addison-Wesley Longman, Boston

Parasuraman R, Barnes M, Cosenzo K (2007) Adaptive automation for human-robot teaming in future command and control systems. The International C2 Journal 1(2):43–68

Reeves B, Nass C (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, New York, NY, USA

Robinson D (1992) Implications of neural networks for how we think about brain function. Behavioral and Brain Science 15:644–655

Stamou G, Vogiatzis D, Stov S (1999) Bridging the gap between subsymbolic and sym-bolic techniques: A pragmatic approach. In: Circuits Systems Communictions and Computers, Athens

Turk M, Pentland A (1991) Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on, pp 586 –591

Van der Ploeg I (1999) The illegal body: 'eurodac' and the politics of biometric identification. Ethics and Information Technology 1:295–302

Van der Ploeg I (2005) Biometrics and the body as information: Normative issues of the socio-technical coding of the body. In: Lyon D (ed) Surveillance as Social Sorting, Routledge, London