



# Augmented Human and Human-Machine Co-evolution: Efficiency and Ethics

Andreas Dengel<sup>1,2(✉)</sup>, Laurence Devillers<sup>3</sup>, and Laura Maria Schaal<sup>1</sup>

<sup>1</sup> DFKI, Kaiserslautern, Germany  
andreas.dengel@dfki.de

<sup>2</sup> University of Kaiserslautern, Kaiserslautern, Germany

<sup>3</sup> Sorbonne-University, Limsi-DATAIA Institute (Saclay), Paris, France

**Abstract.** Capability and expected potential of AI-based computer solutions increased significantly in the recent years, mainly due to progress in machine learning technologies and available data. Growing effectiveness in reasoning, knowledge representation, automatic training via machine learning, and especially in computer vision and speech technology result in AI systems becoming an ever-better communication and work partner of their human counterparts. Deeply embedded in the every-day context of work and leisure, AI systems can act as competent dialog partners and powerful work assistants. Furthermore, they increasingly help humans to better acquire new insights, process and apply situation-specific instructions, receive improved training and learn new knowledge more effectively. Ultimately, intelligent systems exhibit the potential to become inseparable partners of humans or – in case e.g. of prosthesis solutions and innovative sensor technology – even become part of the human body. Such close mental and physical interconnection between human and AI system raises new concerns and ethical questions which need to be considered not only by computer scientists, but ask for interdisciplinary work and social discourse. This paper outlines the different levels of human-computer integration, gives examples of the innovative potential in work assistance and learning support, and sketches ethical and moral issues conjoined with such progress.

**Keywords:** Co-learning · Co-creation · Co-adaption

---

With contributions from: Vincent Alevén, Human-Computer Interaction Institute, Carnegie Mellon University. Kenneth Holstein, Human-Computer Interaction Institute, Carnegie Mellon University. Elisabeth André, Computer Science Department, University of Augsburg. Justine Cassell, Human-Computer Interaction Institute, Carnegie Mellon University. Gordon Cheng, Institute for Cognitive Systems (ICS) Technical University of Munich. Dimosthenis Karatzas, Computer Vision Centre, Universitat Autònoma de Barcelona. Koichi Kise, Dept. of Computer Science and Intelligent Systems, Osaka Prefecture University. Kenji Mase, Graduate School of Informatics, Nagoya University. Raul Rojas, Department of Mathematics and Computer Science, Freie Universität Berlin.

© Springer Nature Switzerland AG 2021

B. Braunschweig and M. Ghallab (Eds.): Reflections on Artificial Intelligence  
for Humanity, LNAI 12600, pp. 203–227, 2021.

[https://doi.org/10.1007/978-3-030-69128-8\\_13](https://doi.org/10.1007/978-3-030-69128-8_13)

## 1 Introduction

Artificial Intelligence (AI) has become more and more important for practical use, especially in recent years, because sufficient computing capacity and correspondingly large amounts of data are available, which especially pushes the evolution of Machine Learning (ML). ML algorithms help people to recognize patterns in existing data sets, make predictions or classify data. Moreover, mathematical models can be used to gain new insights based on these patterns. This holds for many life and business fields, where users often benefit from systems without thinking about the technology in the background. A wide range of ML methods is available for this purpose, including linear regression, instance-based learning, decision tree algorithms, Bayesian statistics, cluster analysis, neural networks, deep learning and methods for dimensional reduction.

The fields of application are manifold and partly known. Think of spam detection, content personalization, such as music and film recommendations, document and sentiment analysis, customer migration prediction, email classification, up-selling opportunities analysis, congestion prediction, genome analysis, medical diagnostics, chat bots and much more. Obviously, there are opportunities for almost all industries and types of companies.

It is a matter of fact, AI plays an increasingly important role as the world becomes more and more complex and poses more and more challenges to individuals, society, companies and institutions. Growing information intensity and information overload, the trend towards shorter innovation cycles and the reduction of knowledge half-live time are all reasons why we face these greater challenges. To penetrate this complexity, AI can make considerable contributions<sup>1</sup>.

There are already impressive technologies for application in professional life that open up new opportunities and potentials: When performing complex tasks, people can fall back on digital companions or use systems that take over entire work packages independently. Such applications are, for example, in practical implementation in the manufacturing industry in quality control or in assembly, maintenance or repair work. Smart applications can also be identified in the field of education [1], such as supporting teaching with the help of intelligent tutoring systems. Sensor data that provide information about eye movements, for example, can help to assess how attentive students are or how well they understand the learning content.

Every technology has its own time and its own impact. AI revolutionizes and permeates our lives in all possible areas. Computers are increasingly taking over the role of a learning partner to enhance performance and productivity, supporting our individual handling of diverse information sources and exploring synergies between large communities. In such an evolutionary cyber-social environment, new potentials for co-creative systems are emerging, assisting users in understanding, learning, decision-making, and memorizing [2]. In the professional world, this is often referred to as Digital Taylorism - a division of labour that brings man and machine into coexistence in order to jointly carry out trial solutions. The term is based on the "Principle of Scientific Management" coined by Frederick Taylor at the beginning of

---

<sup>1</sup> See also chapter 8 of this book.

the 20th century. Breaking-down complex jobs into simple tasks, measuring the outcome of the workers and paying some salary in relation to this outcome is the basic principle of Taylorism. The fundamental axiom of the Digital Taylorism is “what gets measured gets managed”. Thus, the more the technology of measurement advances, the more we hand power to Frederick Taylor's successors.

Today, we have almost unlimited options to measure and this measurement does not only include the classical physical worker but also technicians, managers, and professionals, such as physicists, lawyers, or university professors. Therefore, another way of understanding Digital Taylorism is to describe it as the translation of knowledge work into working knowledge through extraction, codification and digitalization of cognitive tasks into software prescripts that can be solved by AI systems. However, Digital Taylorism does not necessarily mean that people become puppets of digitization. It also does not necessarily mean that people “in the digital world [...] are mere widgets in the giant corporate computer” as described in *The Economist* [3].

Quite the contrary, it is true that AI, if used correctly, with respect to ethical rules, can fruitfully complement and enhance the abilities of humans. With an AI-controlled exoskeleton, for example, a human being can use considerably more power and still implement his sensitive way of performing mechanical actions. When using intelligent systems to perform standardized tasks in the working environment, there is more time to work creatively and apply human problem-solving skills. Correctly used, intelligent tutoring systems can identify and promote the strengths of individuals in school application.

The following article therefore aims to present successful best practices, which were presented in the CoCoLAd Workshop<sup>2</sup> hosted by Andreas Dengel and Laurence Devillers during the Global Forum on AI for Humanity in October 2019. Furthermore, the following examples and statements pursue the objective to raise awareness of the measures necessary for a human-centered co-existence of man and machine in order to achieve a development that is socially and ethically beneficial.

After a short explanation of the terms Augmented Human, Human Machine Co-Evolution (Sect. 2) and approaches for measuring and modelling systems with human-machine interaction (Sect. 3), best practices from the field of education (Sect. 4) are presented. Since the use of such technologies is controversially discussed, also in the field of teaching, this chapter will also outline crucial considerations that should be taken into account when using smart systems. Section 5, the conclusion, focuses on critically reflecting on the presented technologies and giving a short outlook.

## 2 Short Definition of Terms

This chapter will present short definitions of the terms “Augmented Human” and “Human-Machine Co-Evolution” before focusing on the question of ethical principles in AI in general.

---

<sup>2</sup> Human-Machine Co-Creation, Co-Learning and Co-Adaptation.

## 2.1 Augmented Human (Physical/Cognitive/Virtual)

The field of human augmentation focuses on creating cognitive and physical improvements as an integral part of the human body. Let's come back to the already mentioned example of powered exoskeletons: they can improve the quality of life of individuals who have lost the use of their legs by enabling system-assisted walking. While exoskeletons can reduce the stress of manual activity, they may also pose dangers such as potential falls due to a shift in center of gravity.

Advances in artificial intelligence, in conjunction with recent developments in neurotechnology, open the prospect of augmenting and amplifying human cognitive abilities. Neuroscience findings are providing a new level of knowledge for the design of advanced human symbiotic machines that are more tuned to humans. This cognitive augmentation could be beneficial for individuals and society. Cognitive augmentation may be defined as the amplification or extension of core capacities of the mind through enhancement of internal or external information processing systems. Cognition includes acquiring information (perception), selecting (attention), representing (understanding) and retaining (memory) information, and using it to guide behavior (reasoning and coordination of motor outputs). Cognitive stimulation refers to the set of techniques, strategies and materials to improve performance and effectiveness of cognitive capabilities and executive functions such as memory, attention, language, reasoning and planning, among others. Nowadays there are several strategies to train our brain, from classical exercise with conversational agents and serious games to more dynamic, innovative techniques such as brain training games and neurotechnology. In this respect, Sects. 3 and 4 will take up and explain some examples of research topics presented at the CoCoLAd workshop on Human-Machine Co-Creation, Co-Learning and Co-Adaptation.

## 2.2 Human-Machine Co-evolution

People are living together in a “cyber-physical” world with the internet, computers and phones but also cars and connected objects. Smart products have embedded sensors that are continuously connected to the Internet of Things. This applies to buildings and machines, as well as our mobile devices, shopping carts or our sports shoes. The trend is to shift more and more functional intelligence into the products themselves so that they become intelligent agents. This enables them to act independently. Because they are constantly connected to each other via the cloud, whether at home in the four walls, while traveling or at work, and because they synchronize our data with the environment, they can provide us with continuous support. They check their availability, match their skills, coordinate the processing of tasks and control business processes. They also monitor system statuses, optimize material usage, productivity or quality and detect anomalies and redundancies. In doing so, they are constantly learning and adapting to new requirements and changing conditions.

They are thus creating a new form of “simplexity”, in which humans are relieved of the tasks that AI systems can better master. AI thus also becomes a power amplifier technology that complements human skills or enhances their capabilities, both physical and cognitive. The trend is moving away from cooperative assistance systems, through

interdependent human-machine scenarios, to activities where humans and digital agents compete with each other, including in cognitive tasks. The latter applies especially to activities where activity is measurable and understandable. Just as the industrial revolution has neutralized the physical ability of humans in many cases and redefined the division of labour, AI will do the same in the context of intellectually demanding activities and define a new form of division of labour between humans and machines. As a consequence, there is a gradual change in our roles and the roles we give to machines. This way, we may talk about a co-evolution, where intelligent agents and humans mutually adapt to each other through the increasing interaction and interconnection sometimes resulting in an augmented human.

The interactions with intelligent agents, conversational robots are already a kind of enhancement technologies. In order to augment our performances, computers and robots are also increasingly taking over the role of a learning partner. The capabilities of emerging technologies are underpinning the formation of new human-machine partnerships, which will have significant impact on both individuals and organizations. More specifically, these human-machine partnerships<sup>3</sup> have the potential to allow people to find information and act on it without emotional interference or external bias, while exercising human judgment where appropriate. If we learn to “team up” with technologies integrated with human-machine learning tools, we can imagine a future in which this collaboration helps provide the resources and knowledge we need to manage our daily lives.

Recently, the research focus in the field has moved to mobile and pervasive interaction, including embodied interfaces and intelligent user inter-faces. However, most of the time, there is still a clear separation between the user and the system. The augmented human of the 21st century with physical exoskeleton, bionic eyes or prostheses, cognitive stimulation or virtual experiments fascinates and repels us at the same time. Where should the red line between repair, care and augmentation actually be drawn?

Designing and developing great AI systems that allow users to effectively interact or work together is no easy task. If you google “the C’s of social technology Interaction” you will get links to a myriad of “C-words” including: Collaboration, Communication, Cooperation, Creativity, Coordination, Critical Thinking, etc.”. All of which are important elements of learning and working and can be enhanced with the use of technology. In order to describe the interaction between humans and robots working together, three scenarios have been established in the professional world: Coexistence, Cooperation and Collaboration. In the coexistence scenario, humans and robots work in separate workspaces, with no interaction or overlap between humans and robots. CoBot is the contraction of “collaborative” and “robot”, name and concept of a new kind of robots able to work literally hand-in-hand with humans without a safety fence between them. In the cooperative scenario, humans and robots work simultaneously in the same workspace on different objects or tasks. In the collaboration scenario, man and robot work hand in hand on a common task or object. The robot assists humans, for example, when adding components to be assembled. The AI

---

<sup>3</sup> <https://www.iftf.org/humanmachinepartnerships/>.

systems that will be most useful to us in the future are those that collaborate rather than replacing, those that cooperate rather than competing and those that can effectively co-exist with humans. Going from human-robot coexistence to collaboration is a real technological and social challenge.

### 2.3 Core Principles for Ethical AI

Designing and developing great collaborations with AI systems that respect ethical principles is no easy task. For example, emerging interactive and adaptive systems using sophisticated skills like emotion detection or simulation [4] modify how we will socialize with machines with positive impacts but also some risks. On the one hand capturing, transmitting and mimicking our feelings will open up new applications and better collaborations with machines in health, education, transport and entertainment. On the other hand, these areas inspire critical questions centering on the ethics, the goals and the deployment of innovative products that can change our lives and society. Such close mental and physical interconnections between humans and AI systems raise new concerns and ethical questions which need to be considered not only by computer scientists, but through interdisciplinary work and social discourse regarding the different areas of application.

Several high-profile initiatives established in the interest of socially beneficial AI have been proposed. A unified framework may therefore be synthesized [5] from these approaches which tries to define goals and limits of AI systems and their development, consisting of five core principles for ethical AI:

- Beneficence: promoting well-being, preserving dignity, and sustaining the planet
- Non-maleficence: privacy, security and “capability caution”
- Autonomy: the power to decide
- Justice: promoting prosperity and preserving solidarity
- Transparency and Explicability: enabling the other principles through intelligibility and accountability

Ethical issues must be treated in more depth for each application. The use of AI in education, health, etc. will bring great benefits if we can audit the systems and verify these core principles for ethical AI<sup>4</sup>.

## 3 Facets of Human Machine Co-creation, Co-learning and Co-adaption

The integration of cyberspace with the real world, which is called “cyber-physical world” or “digital twin” today, is rapidly advancing based on improvements in AI, robotics, data analytics, virtual reality and the internet of things, which are penetrating our society. People are working and living together in such cyber-physical world. Since we interact with robots and smart agents or use machine-assistance, our living style,

---

<sup>4</sup> See also chapters 10 and 11.

performance and functions are already being assisted or augmented by these technologies. Oftentimes, the systems we interact with act “human-like” or perform human tasks. The following subsections will therefore present three concrete methods of modelling how machines can learn from humans (Sects. 3.1, 3.2 and 3.3), before finally focusing on the question of how to design symbiotic society envisioning proper and human-beneficial cyber technologies in general (Sects. 3.4 and 3.5).

### 3.1 Surviving in Man-Made Environments: The Case for Language and Vision<sup>5</sup>

It is easy to imagine a future where social, intelligent machines interact with humans and can successfully complete everyday tasks for or with us, such as doing our shopping, or helping us getting around the city. In such a scenario, it is inconceivable to imagine machines, aimed to co-exist with humans in man-made environments, that are not able to understand and use language, be it written or verbal. Language is a key instrument of human intelligence – intimately linked with vision. Our visual interpretation capacity is jointly acquired with the linguistic structures we use to describe the world. As such, it makes sense to address the acquisition of vision and linguistic skills by machines jointly, as complementary facets of machine cognition.

Computer vision, reading systems and natural language processing have been key and challenging<sup>6</sup> research areas of artificial intelligence and have independently advanced for many decades. Ultimately, the research community has started to explore the interconnections between them. It is quite plausible that future machines will learn to interpret images and language jointly, in a multi-modal fashion, like humans do. And of course, they will be using natural language to interface with humans. The first skill we would like machines to possess is the capacity to read written information in the world around us.

Text is omnipresent around us, especially in urban environments. Importantly, when text is present, it usually carries high-level semantic information, vital to fully understand the scene. Until very recently, the computer vision community has ignored text appearing in real scenes. Nowadays, various researchers work on multiple topics related to reading text in the wild [6], from large-scale text spotting and scene-text based image retrieval [7], to end-to-end reading systems for specific applications [8]. An important tractor for recent advances has been the Robust Reading Competition series, which has consistently pushed the community forward by proposing new challenges and scenarios (from multi-lingual [9] to driving [10]) and offering a consistent evaluation framework.

Following numerous years of research in this field, it has become obvious that reading text around us is not an end on its own, but makes more sense in the context of

---

<sup>5</sup> A contribution based on the GFAIH-speech and research work of Dimosthenis Karatzas.

<sup>6</sup> As well described by Hans Moravec’ paradox, seemingly easy aspects of intelligence involving perceptual and motor skills appear easy to humans who count with a long evolution process mastering these skills, but tend to be much more challenging to solve and engineer than higher-level aspects of intelligence such as reasoning which might appear perplexing, but are not intrinsically so difficult to achieve.

interpreting the scene as a whole. How does textual information relate to the visual aspects of a scene, and vice-versa, what can a quick glimpse of a scene tell us about the textual content we expect to encounter there? It turns out that there are many different ways we can learn to associate visual content to textual context (see e.g. [11]). For example, it was shown how peeking at a scene can optimize the subsequent text recognition processes by producing contextualized language models [12] that reflect the “topic” of the image.

Scientists could also demonstrate, that in the process of jointly learning the visual and textual modality, joint representations that effectively map an image to a semantic space defined by the text were learned.

Indeed, it was shown how semantic representations can be learned by feeding the whole of Wikipedia to a neural network model and forcing it to predict for each image what topic (as expressed by the linguistic content of the associated article) it could be used to illustrate [13]. This joint modelling of vision and language has many applications apart from self-supervised learning [13], from cross-modal retrieval [14], to fine-grained classification [15] or hate speech detection in social media [16].

A natural extension of these ideas is exploring the links between vision and understanding or producing natural language. People understand scenes by building causal models and employing them to compose stories that explain their perceptual observations [17]. This capacity of humans is associated with intelligent behaviour. The ability to describe an image is one of the oldest cognitive tasks in intelligence tests [18], and it is intimately related with our capacity to build and employ such a causal model to explain the world.

Current state of the art image captioning models (e.g. [19]), still behave like 5-year olds, enumerating objects and at best describing their visual appearance and relative positions, keeping short from actually interpreting the scene, and producing plausible explanations for the depicted content. In this sense there are recent advances aiming to shifting captioning models towards producing image interpretations, by incorporating prior world knowledge to the visual analysis of the image. What is even more interesting is a bi-directional interaction between human and machine. Imagine a blind person asking an intelligent agent what temperature the air-conditioning is set at, or whether a can of beans has expired. These are real-life questions asked in this community [20], which the AI researchers currently have no way to deal with. Being able to ask a question about the world using natural language, that an intelligent agent is able to understand and respond to in natural language, by combining visual and textual information in the scene, in a fully multilingual setting, is probably one of the best scenarios to drive progress and bring vision and language research together [21].

Both computer vision and natural language processing are data-driven disciplines. As such, it is well known that the resulting models suffer from biases derived from the data used to train them. For example, gender bias is a known problem of captioning systems. Many systems would be more probable to suggest that the person seen in the scene is a man when a skateboard or wind surf action is depicted [22]. Of course, the



problem does not stem from the model or the learning algorithm, but instead from the data, and ultimately the society that generated them<sup>7</sup>. The fact is that annotators shown an image of a skateboard are more probable to describe it as a “man” than as a “woman” riding the skateboard. This reflects our own biases, and it would be unfair to blame the captioning model for the shortcomings of our own society.

Unfortunately, the media is usually fast to blame the learnt models, and AI as a whole, for these shortcomings. In many ways this is a “shooting the messenger” reaction, blaming the data-driven models for bearing the news that our society is indeed rid of biases of all sorts. In reality, researchers are actively looking into ways to compensate for data bias [23, 24].

The evaluation metrics used to measure the performance of vision and language models are also a source of worry. Usually, the performance of visual question answering is measured just by the accuracy of selecting the right response, leading models to learn typical correlations between questions and answers instead of really understanding the image. Similarly, captioning systems are measured by the degree by which resulting sentences match a set of human produced captions, resulting to models that can easily reproduce typical linguistic structures, but cannot describe anything slightly unusual.

Measuring performance is not trivial when it comes to such high-level tasks. Recent works on a system for producing captions of newspaper images, using the associated article as a source of contextual information [25] led to a system able to produce plausible captions, describing the people and places in the image. It is not possible to judge the quality of such results just by comparing them to the original caption of the professional journalist. But most importantly, it is impossible to automatically measure the correctness of such captions – many of the captions will appear plausible, while the model might attach the wrong name to a person or a location, leading essentially to problems in detecting “fake captions” if we only bother about using standard evaluation metrics. Human-in-the-loop methods, complementing automatic evaluation, are extremely important in this space.

### 3.2 Robots Learning from Humans: Past, Current and Future to Purposive Learning<sup>8</sup>

In the workplaces of the future, people will be able to perform complex tasks with the help of digital companions who can see, hear and touch and thus perceive their surroundings. Communication and interaction with information and physical objects will be facilitated by personalized support adapted to the context of the task, the environment or the performance, and tailored to individual workplaces. This specifically holds for human-robot interaction.

In order to train robots to support humans, there are various options, one of them is Purposive Learning. This method, as it has been pointed out, reasoning about the meanings of observed human activities, is a powerful way for robots to learn from humans, and learning from humans is a powerful means to ensure meaningful human-

---

<sup>7</sup> See also chapter 6.

<sup>8</sup> A contribution based on the GFAIH-speech and research work of Gordon Cheng.

centric outcomes [26]. Fundamental studies in human imitation learning have revealed that behavioral imitation is the central aspect of cognitive development in humans. Essentially, it has been noticed that a simple direct copy of observed movements has little meaning, this is due to the different embodiment of the imitator, which does not normally match the embodiment of the observed demonstrator. One of the earliest seminal works in robot imitation learning was by Kuniyoshi et al. [27], which showed that it is essential to extract specific features that match the demonstrator and the imitator at the start of the imitation process. Based on human sciences studies, three levels in imitation learning (see Fig. 1) were derived:

- **Appearance-based:** at this level, the imitator usually focusses on the reproduction of the motion of the demonstrator
- **Action-based:** at this level, the imitator will focus to select an action based on already known actions in an attempt to closely match the observed demonstration
- **Purposive-based:** focusing on the intention/goals of the entire observed task, that is to extract a deeper understanding of the observation

Appearance-based strategy is the most common approach in robotics. Dynamics Movements Primitives (DMP) is a well-accepted method used by the robotics community, as it can generate and encode trajectories in an adaptive form [28]. Whereas, the Action-based strategy requires to learn a correct mapping between the action and the capability of the robot.

To ensure the success of this strategy, a policy is learned as to what and when to perform the particular action by the robot. Earlier works in this area showed results that a robot can deal with very dynamic situations. For instance, learning to play a game of air-hockey [29].

Roboticians usually focus on the realization of a single task that is fairly fixed in an environment with little variances within the task. Thereby, limiting the scale of the task's complexity and making its difficult of generalization into other domains.

Purposive-based learning set out to tackle the core issues of generalizable learning to enable robots to learn from humans in a more flexible manner [26]. Thus, enable robots to reason about the meanings of human activities. This approach is considered as a powerful way for robots to learn from humans based on the answering fundamental questions on: How can we move beyond the learning of single tasks and ensure that generalizable human observations can be reused across multiple tasks and domains?

The new novel learning approach that utilizes artificial intelligence (AI) methods for inferring semantics with reasoning methods, such technique has been able to induce two fundamental changes: i) extracting semantic (meaningful) representations from the human behaviors from observations; and ii) the ability to transfer and/or reuse past knowledges in new domains. Furthermore, these AI methods have shown to produce a compact and them human-readable representations. Furthermore, it has been shown that the prior knowledge can even enhance low-level perception [26, 30]. Knowledge-based representation can provide us with a powerful mechanism in dealing with invariance, thus, yielding reusable and generalizable knowledge [31]. Such works have shown that even complex observations can be dealt with, such successfully learning from observing multiple humans' performance of the same task in different styles [26, 30].

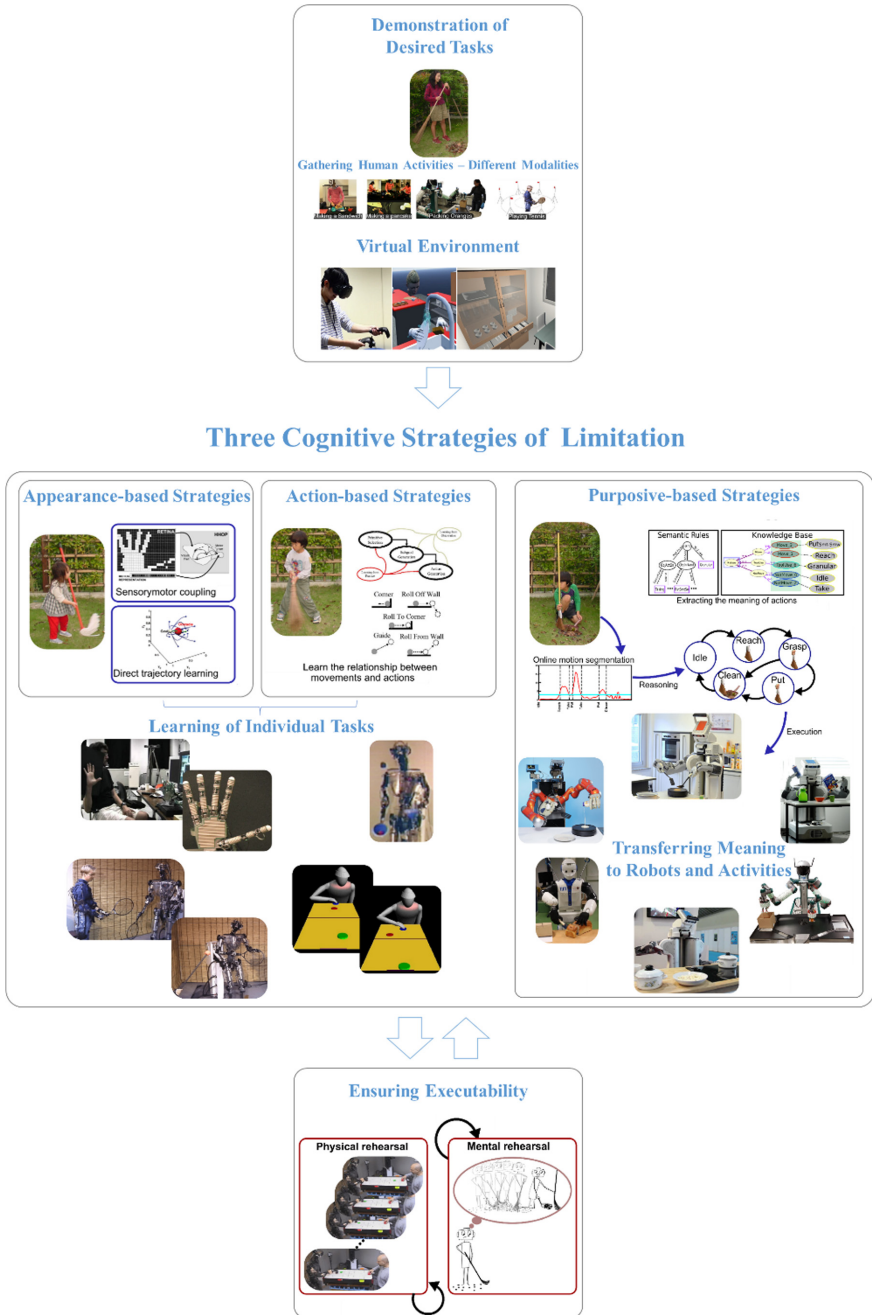


Fig. 1. Purposive Robots Learning from Humans: Overview of three strategies [26]

### 3.3 Empowering Multimodal Affective Behavior Analysis by Interactive Machine Learning<sup>9</sup>

Another facet that needs to be considered in the process of modeling human behavior is the description of human affective behavior. Well described corpora that are rich of human affective behavior are needed in a number of disciplines, such as Affective Health Monitoring or Behavioral Psychology. However, populating captured human behavioral data with adequate descriptions can be an extremely exhausting and time-consuming task. Therefore, attempts are being made to facilitate the acquisition of annotated data sets by involving end users directly in the Machine Learning (ML) process.

Users are enabled to interactively enhance their ML model by incrementally adding new data to the training set, while at the same time getting a better understanding of the capabilities of their model. In the approach presented in [32] and [33], this happens on multiple levels. First, users get a pure intuition of how well their model performs, by investigating false predicted labels. They may even learn specific cases in the data when their model “always fails” or when they can be sure they can trust their model. Secondly, besides intuition, so-called explainable AI algorithms provided within the workflow allow users to generate local posthoc explanations on instances their model predicted. This way interactive ML techniques and explainable AI algorithms are combined to involve the human in the ML process, while at the same time giving back control and transparency to users. In that sense a combination of three recent topics of ML takes place:

- Explainable Artificial Intelligence, as the transparency of the decision process is increased via visualization of the predictions
- Semi-Supervised Active Learning, since labels with low confidence are highlighted to guide the user towards relevant parts
- Interactive ML, because human intelligence and machine power can cooperate and improve each other.

The overall approach can be subsumed under the term eXplainable Cooperative Machine Learning (XCML). Researchers in this field strongly believe that disciplines such as health care, psychotherapy, and others may benefit from XCML technologies. Especially in high risk environments that apply artificial intelligence it is crucial to not only rely on high prediction accuracies, but also to fully understand the underlying processes that led to a classification result (see also the criterion “Transparency and Explicability” of the five core principles of ethical AI) For further information see references [32] and [33].

---

<sup>9</sup> A contribution based on the GFAIH-speech and research work of Elisabeth André.

### 3.4 Symbiotic Interaction to Socialware – Social and Semantic Interactions of Augmented Human and Ambient Intelligence<sup>10</sup>

Having the examples of the last three subsections in mind, it is easy to understand that the game is changing in many areas. Therefore, the research field has to be extended as recent technologies are showing us a future vision of realizing smart information environments and augmentation of human abilities. The aim has to focus on creating and developing core information technologies that realize advanced interaction designs for a symbiotic society consisting of humans, augmented humans, connected things, ambient intelligence (i.e., a smart intelligence environment), internet of wisdoms, robots, etc. Such advanced interaction in the symbiotic society can be called “Symbiotic Interaction.”

Researchers in Japan Science and Technology Agency (JST) CREST program on Symbiotic Interaction area aim to create and develop the fundamental technologies that realize symbiotic interaction based on understanding and designing interactions in a symbiotic society. The goal of this research area is to establish core technologies of symbiotic interactions through approaches that evaluate behaviors of humans and societies, designing future societies, and constructing effective interactive systems. It covers state-of-the-art technologies in appropriate areas such as human-computer interaction, ubiquitous/wearable information processing, computer science, and robotics, in addition to collaboration with other disciplines such as cognitive science, social science, and brain science. For examples, there are projects on tender elderly care skill training technology to promote well-being, humanoid robotics to enlighten moral in public space, and speech synthesis and recognition technology for secure and spoof-free speech-based services and protection of privacy, etc. Computer vision technology of human behavior and interaction and wearable IoT devices support and utilize analytics of staring gaze and touch interactions during care practice (Co-Learning). So-called “*Moral robots*” will cooperate with human to create secure and comfortable public space and retail business (Co-Creation). Spoof-free and realistic speech synthesis technologies will lead the deep discussion of relationship and utilization of advanced AI technology and personae (Co-Adaption).

Following these aspects, research and development efforts will contribute to establishing a harmonized, human-centered and globally-optimized symbiotic society that benefits by rapidly advancing AI technologies and fundamentals.

The computer architecture today is well-known as the stack of hardware and software on it. With the symbiotic interaction research, the social interaction parts are put together and form a novel architecture of platform for symbiotic society.

Socialware contains traditional context processing, semantic processing with interaction data at signal processing and machine learning tools. Within the Socialware, knowledge base, inference and ontology technologies are incorporated to construct symbiotic interaction corpus and dictionary, which will be used as a basic common sense of robots and intelligent systems. They are most useful for robot and intelligent systems to co-work and assist flexibly with variety of humans. Cognitive human

---

<sup>10</sup> A contribution based on the GFAlH-speech and research work of Kenji Mase.

models in symbiotic society and its social design principles should be included in the Socialware, too. Socialware plays the role of foundation of important applications/innovations in the symbiotic society of the digital twin.

### 3.5 Socially Aware AI - Maintaining the Human at the Center of AI Design<sup>11</sup>

Another model also concentrates on the social aspects of human interaction, and the need to consider them in the design of AI systems. This model starts at a different point, however. It focuses on a development methodology that takes into account from the beginning of the design process the importance of designing systems capable of co-adaptation –the dyadic processes whereby people and AIs adapt to one another in real time. It also relies on the perspective of conversation as co-created by two (or more) interlocutors. And intrinsic to the design methodology is attention to ethics – an attention to what systems we decide to design, and in what order, based on the grand societal challenges of the day.

The model of “Socially aware AI” stems from the fact that somewhere along the path of defining and shaping AI, as we have been doing since the 50s, the definition of AI itself has changed. Today most researchers have abandoned the goal of simulating human intelligence. Instead, they wish to build systems that can do what humans do, only better. Systems that can read X-Rays of human lungs, but with a higher accuracy rate than doctors. Systems that can understand human speech, better even than humans can. These systems emulate human intelligence and human abilities. Problems may arise due to the fact that no roadmap exists to describe which human abilities should be emulated first – and which should never be emulated. Therefore, the question has to be asked: What should AI systems be designed to do, and what should we prevent them from doing? One answer is to ensure that the design process be guided by the following human-centered principles:

- The principle of the “3 Cs”: Coexistence, Cooperation, and Collaboration. The AI systems that will be most useful to us in the future are those that **collaborate** (rather than replacing), those that **cooperate** (rather than competing) and those that can effectively **co-exist** with humans
- The principle of urgent societal need: Grand societal challenges must be addressed first– such as inequity, illness and disability, poverty.<sup>12</sup>

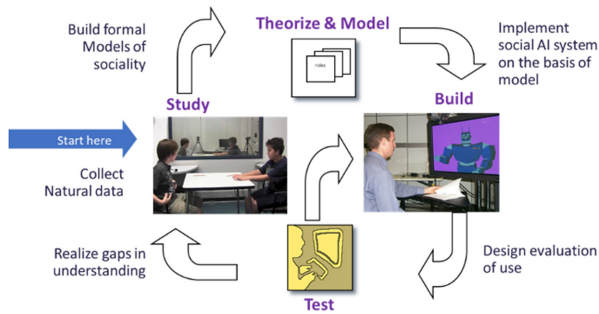
This view leads to the socially aware AI Methodology depicted in Fig. 2, that can be used in the development process. This approach can be called **Socially-Aware AI** [34, 35] as it is socially-aware in two ways:

1. In addition to being able to effectively carry out a task, the system is aware of social norms and abilities, and is able to use them to more effectively work with people;
2. In addition to innovating technically, the designer of the system is aware of tough social problems, and is dedicated to addressing them.

<sup>11</sup> A contribution based on the GFAIH-speech and research work of Justine Cassell.

<sup>12</sup> See also chapter 8.

These principles seem straightforward - however, few AI researchers stop to think about what an AI system needs to know in order to cooperate or collaborate. Nor do many stop to look around and ask what grand societal challenges need to be addressed. These concepts of social awareness imply human abilities that have rarely been modeled in machines – the ability to get along, to build a bond, to inspire trust, to listen well. Systems of this sort need to know how to amplify human abilities, as well as to have strong abilities of their own. Rather than manipulating human behavior, Socially Aware AI inspires learning about oneself. Rather than trying to make the most humanlike chatbot, Socially-Aware AI targets just enough human-like behavior to bring out the best in its human partners.



**Fig. 2.** Socially-Aware AI methodology

Results of existing Socially-Aware AI systems have been able to achieve groundbreaking results: They have effectively taught children with autism how to build social bonds with their peers [36]. They have inspired world leaders to reveal their likes and dislikes so that the system can better assist them [37]. And they have inspired social bonds strong enough to lead to stronger science learning in children in educationally impoverished neighborhoods [38]. A particularly poignant example is the Alex Virtual Peer project. A virtual peer is a cartoon life-size virtual child on a screen. Results of this work have shown that a virtual peer that speaks the same marginalized dialect as a child is capable of inspiring increased rapport – a close bond – with that child, and that rapport between child and virtual child predicts improvement in the use of classroom science talk [38]. Marginalized dialects include African American English, Verlan-influenced French and Newcastle UK English, among many others. They exist in all countries. They are often thought to be signs of poor education, when in fact they are simply separate linguistic varieties. Teachers do not necessarily speak these varieties, but putting AI-based virtual peers in the classroom that do speak like the children can therefore improve the classroom performance of children from marginalized communities. These are the kinds of societal grand challenges that social-aware AI can address.

## 4 Best Practices in Education

Whenever AI is linked to the school system, alarm bells go off for many people. Monitoring systems are prematurely imagined that collect data about pupils that go far beyond their meaningful use in class. The fear of pupils becoming “transparent” through surveillance is growing, combined with the fear that data about these students could be misused and deployed to evaluate other areas of their lives. However, the examples in the following Sects. 4.1, 4.2, 4.3 show that the use of AI in the educational sector does not need to mean that teachers are ousted and children are henceforth taught by self-sufficient AI systems that collect an inappropriate amount of data. In fact, systems that support teachers, and that take on tasks that may be difficult for teachers to deal with (such as speaking in the same dialect as the child), can also have a big effect. Systems such as these, however, can only be implemented if the design phase includes a careful observation and understanding of children’s lives – a truly human-centered approach.

The goal of any sensor-based detection, the collection of data, should always follow the goal of using this data to determine what kind of content is effective for which learners and how. It is then the task of didactics to develop appropriate materials in order to provide individually tailored educational measures or to better challenge and promote individual learning needs and ultimately also to be able to measure the effects.

### 4.1 IntelliChalk – Teaching Mathematics with a Data Wall<sup>13</sup>

The changes mentioned in the introduction do not stop at the education system. Previous forms of teaching must face up to our dynamic times and, ideally, overcome traditional forms of learning and the use of media. Intellichalk, which means “intelligent chalk (board)”, is an innovative way to design today’s teaching. At the *Freie Universität Berlin* a large data wall composed of computer screens has been used teaching mathematics and natural sciences (Fig. 3). The idea is to apply the three C’s mentioned in the introduction: the digital chalkboard **collaborates** with the lecturer, **cooperates** providing assistance, and **co-exists** with humans. It is not aimed at making the lecturer superfluous.

In comparison to traditional chalkboards, which are normally used in schools or universities, the contrast of the digital screens provides a much better visual experience. Students sitting in the last row can still see the diagrams and formulas clearly. The lecturer writes on a contact sensitive tablet which offers several functions: It is a drawing program which provides the lecturer the tools to draw and write with high quality as well as its a program which manages images for pasting, as well as scans of handwritten notes.

The developers of IntelliChalk keep on improving the software to include more features such as for example slide presentations via IntelliChalk or interactive lecturing using a contact sensitive screen mounted on a podium. Furthermore, handwriting recognition can be used to start secondary applications such as simulators, algebraic servers, or an image search over the Internet as well as videos can be pasted to the board.

<sup>13</sup> A contribution based on the GFAIH-speech and research work of Raúl Rojas.





**Fig. 3.** The podium for the lecturer includes a contact sensitive screen

Lectures are available over the Internet as a file for printing, or as a file for replaying the lecture. Handwritten notes of the lecturer can be digitized in a few seconds before the lecture starts and students themselves can annotate their own local copy of the class material using their own tablets. In this way, the student's annotations constitute an additional information layer.

The developers of IntelliChalk think, that this will be the future of teaching on site or via conference mode in universities and also in schools. The system can be imagined as an AI that co-creates the lecture, by providing, for example algebraic processing and simulations on demand. The system can become better over time, co-learning from previous lectures and the materials produced.

#### **4.2 Lumilo – AI for Personalized Learning: Students, Teachers and AI Systems Augmenting Each Other's Abilities<sup>14</sup>**

The example of Lumilo addresses a real-time, mixed-reality teacher support tool. It is an instance of human-AI complementarity in the domain of mathematics instruction. Lumilo augments teachers' in-the-moment decision-making regarding how best to help their students. It is a result of the dissertation research of Kenneth Holstein at Carnegie Mellon University, in the Human-Computer Interaction Institute.

Many applications of AI may be most effective when designed from the start to be synergistic with human intelligence. To achieve such synergy, designers must deeply understand how, in the given task domain, humans and AI can augment each other, based on their complementary strengths and weaknesses. Human-centered design practices have much to offer in this regard, since they center human needs and abilities in the design process. However, prototyping novel human-AI interactions is still a relatively new challenge, requiring innovation in design methods and processes.

<sup>14</sup> A contribution based on the GFAIH-speech and research work of Vincent Aleven and Kenneth Holstein.

Lumilo was designed with this knowledge in mind and was specially tailored to meet the challenges mentioned above. It is designed to help teachers dynamically prioritize which students may need teacher attention, as a class of students works with AI-based tutoring software, an increasingly common scenario in schools in the US and elsewhere. The mixed-reality tool projects, in the teacher's view of the classroom, an indicator of each student's progress or struggle. "Deep Dive" screens in Lumilo provide teachers with more detailed information about a student to provide more context as needed, to aid teachers in deciding whether and how to help a given student.

Lumilo was created over a period of two years, during which its developers worked extensively with middle school teachers. A variety of methods of human-centered design were employed to gain a deep understanding of their needs, strengths, and boundaries, and of how best to take advantage of the many existing learning analytics developed over two decades by the fields of AI in Education, Learning Analytics, and Educational Data Mining. Through many rounds of iterative prototyping, the tool was honed for classroom use, based on extensive teacher feedback. In the process, new methods for human-centered design were developed, namely, a new prototyping method for dynamic data-driven AI algorithmic experiences, called Replay Enactments, and a new method for the iterative, evidence-centered design of teacher-facing analytics tools, called Causal Alignment Analysis.

The effects of Lumilo were tested in a classroom study with 286 middle school students, across 18 classrooms and 8 teachers. All students used AI-based tutoring software for 2 class sessions in order to hone their skill in equation solving. Classes were randomly assigned to conditions which differed only in whether the teacher used Lumilo or not<sup>15</sup>. Teachers using Lumilo were guided by Lumilo's mixed-reality indicators and Deep Dive screen in their decisions of whom to help, and how. Without Lumilo, teachers had to rely on their own observations and judgment to decide which students to help. This condition represents business-as-usual in classes using intelligent tutoring software. Results show that teachers, when using Lumilo, devote measurably more time to students who have more to learn (as compared to other students) than they do without the tool. As a result, students learn more, especially those who had more to learn. Interestingly, in the Lumilo condition, pre-test scores were less predictive of post-test scores than in the other conditions. Thus, Lumilo helps teachers enact more equitable practices in classrooms, where students who have more to learn get more attention and have greater learning gains.

The work illustrates the creation of an effective new human-AI partnership through human-centered design. The AI augments what teachers do: The teachers we observed do not defer to the AI; rather, they interpret Lumilo's indicators and Deep Dive screens against what they glean from observing the classroom and what they know about their students. Demonstrations of successful human-AI partnership are rare, especially for complex tasks carried out in authentic, real-world settings. The work illustrates that careful use of human-centered design processes can be highly effective to this end, and

<sup>15</sup> In a third condition, teachers used an ablated version of Lumilo, to evaluate the added value of the analytics within Lumilo over and above other elements of Lumilo's design. For brevity, this comparison was omitted.

illustrates as well that new methods may be needed to design for human-AI synergy. Like the developers of IntelliChalk, Lumilo developers also see the future of new teaching in these applications and anticipate that many novel methods will sprout up in the nearby future. For further readings see [39, 40] and [41].

#### 4.3 Wordometer, CoaLA and LeAE – Experiential Supplements: Sharing Human Experiences for Co-learning<sup>16</sup>

Experiential supplements are pieces of information extracted from human experience and employed to help humans to solve their problems. This concept of utilizing human experiences is based on the observation that humans continue to face problems that have already been solved by other humans. In the context of learning, a learner can help other learners by sharing his/her experience of overcoming the problem he/she has already faced. Computers can help co-learning among learners by providing the mechanism to share learners' experiences. AI technologies that sense and estimate learners' current knowledge levels, mental and cognitive states play important roles for experiential supplements. Another important role of AI is how to produce and apply experiential supplements. Generally speaking, learners react differently to the same information. In other words, we need to prepare prescriptions of experiential supplements: to whom and when an experiential supplement should be applied to improve learner's states.

The notion of experiential supplements is to build a computer system that assist humans to help others through sharing experiences. Co-learning among humans is implemented by the system. In this sense, the system realizes intelligence augmentation or "inclusiveness" of AI. In the context of learning and similar to the aspect of cognition augmentation mentioned in Sect. 2.1, one can call it "learning augmentation": an AI system helps a learner learn better.

Because the system works in a fully person-dependent way, we need to be careful about the "fairness" to learners. An experiential supplement can be different for learners having the same problem. A learner may complain that his/her problem cannot be solved due to a different experiential supplement given to him/her. Thus, accountability is also an important ethical aspect. Human experiences are personal in nature. Thus, privacy is also significant in this framework. In particular, the right of persons who provide experiences must be protected.

In the following three systems, Wordometer, CoaLA, and LeAE, experiential supplements for learning, which aim to improve learning by using other person's learning experiences, are presented:

**The Wordometer** is an application by which the total number of read words in a certain period (typically in a day [42]) is measured. Based on this approach in [43], we presented four nudging strategies for sustaining or improving user's engagement of reading documents: showing the number of read words, setting up the goal of reading amount, notification of typical locations and timing of reading, and sharing the number of words with a peer group. Setting up the goal and using the peer group are the

---

<sup>16</sup> A contribution based on the GFaiH-speech and research work of Koichi Kise.

nudging strategies worked well to improve the amount of reading. Machine learning is employed to build a prescription for each nudging strategy. By taking into account the personal traits of a learner, the system can select appropriate nudging strategies to help the learner. More information may be found in [43].

**CoaLA** is a system for confidence-aware learning assist. It is capable of estimating the user's confidence in his/her answer to a question. It uses an eye-tracker for the estimation because eye movement reflects the user's internal states such as confidence [43, 44]. Given eye movement data as input, it is possible to estimate learner's confidence by using machine learning. Based on the estimated confidence, cases of correct answers without confidence (correct answers by chance), as well as incorrect answers with confidence (misunderstanding) can be detected. By notifying them to the use, the quality of knowledge has been successfully improved.

**LeAE** stands for learning with an aerobic exercise. This enables us to memorize new words better with the help of an aerobic exercise, using a stepper. The experimental results have shown that the number of remembered words is larger after three days and one week, in the case that words were memorized with the aerobic exercise. The difference between with and without the aerobic exercise was statistically significant ( $p < 0.01$ ). However, the aerobic exercise worsened the performance for some users. Thus, it is necessary for us to build a prescription to distinguish users with positive effects from those without them.

Currently, systems with a more advanced way of co-learning are already in the making. In the above examples, the system has learned prior to its application. However, due to the lack of training data, the learning itself is a difficult task and recent work therefore concentrates on working on the co-learning of the learner and AI. AI can learn from the behavior of the learner for better estimation of his/her internal states, as well as strategies of human learning. On the other hand, the learner can learn from the learned AI which can provide a fully personalized strategies of learning. A possible scenario is the adaptive generation of exercises by AI to maximize the learning effect as well as motivation of the learner.

## 5 Conclusion

The examples of the last chapters describe anything but horror scenarios that can be imagined in the context of AI and its applications. We discussed a variety of different topics such as intelligent vision and language models and robots learning from humans, socially aware AI or best practices of smart systems and applications in education. What all examples have in common is that we humans play an important role. We must take responsibility for these systems and it is important not to ignore the fears and dark sides that technology can bring and to remain sensible to the important questions that have to be asked. The successes of AI in recent years and the applications of Augmented human and Human-Machine co-evolution have led to much speculation about the capabilities of these technologies that must be clarified.

In general, positive impacts might be seen through the development of human-centered AI, aware of social norms and abilities, and the capacity to efficiently improve work with people. Risks might be dependency, isolation, dehumanization and

manipulation especially for vulnerable people. In the case of systems that imitate human emotions, problems can arise especially in the interpretation of these emotions and in the classification of these machines in our society. The recommendations are to clarify the limits of imitation to avoid over attribution of capacities and to keep a clear distinction between a living being and a machine [45]. Another important point to note is that systems change when they continue to learn after deployment. Who is responsible if the machine malfunctions: the designer, the owner of the data, the owner of the system, its user? [46]. The machine itself cannot be responsible. Users should be aware of the learning capacity of the machine that can lead to new issues that affect the consent of both user and society. Because long-term behavior is difficult to control, machines should be controlled with benchmarks several times during the time of usage. Researchers should seek to contribute to societal debates and to the development of assessment benchmarks and protocols for broad dissemination of machine learning systems. For use in specialized professional sectors (medicine, law, transportation, energy, etc.), data collection and analysis require collaboration between computer scientists and experts in those fields.

In summary, one can therefore state, that regarding the discussed technologies, in view of the strategic stakes as well as the impact on the economy and society, the scientific aspect alone is not enough. It is also necessary to examine the ethical and societal issues raised by the development and deployment of AI independent from its application field, and to propose concrete frameworks to address them. The shown examples also illustrate that there is an awareness of the need for action and researchers are trying to find solutions to ethical problems. Events like the GFAIH also contribute to this by catalyzing interdisciplinary exchange and generating recommendations for action. These go beyond their application in science and must also be communicated to the people who use such technologies. It is important to demystify and disseminate AI science whether it is used in terms of a learning partner, a digital assistant in a factory or as a robot: Imagination of our contemporaries about robotics and more generally AI are mainly founded on science-fiction narratives and myths. Expressions used by experts such as “robots are autonomous”, “they make decisions”, “they learn by themselves” are not understood as metaphors by those outside the technical research community. To mitigate ideas originating from science fiction that mainly underline gloomy consequences, it is important to engage in public discussion and debate with all citizens.

Emerging technologies proceed through multiple stages of evolution: from early stage research, experimentation, prototypes, testing, validation, evaluation and societal adoption. The ethical considerations can be analyzed at each stage of development. Researchers must also ask themselves about the usefulness and the effects of the artificial and the natural of the resemblance to the living and take care to communicate this clearly to the public.

An Observatory on Society and Artificial Intelligence (OSAI) has been also created in Europe. It aims at offering a set of tools that help people better understand and study the impact of AI technologies across the European Union. Specifically, the Observatory supports the distribution and the discussion of knowledge about the Ethical, Legal, Social, Economic and Cultural issues of AI (ELSEC-AI) within Europe. We must amplify these initiatives all around the world and share the results.

Without safeguards against the deployment of products capable of manipulating our emotions and decisions, continuously present in our intimacy, we would be playing sorcerer's apprentice. The development of AI is a business, and businesses are notoriously not interested in fundamental ethical guarantees. The Global Partnership on AI, (GPAI) which is an international, multi-stakeholder initiative to guide the responsible development and use of AI, in a spirit of respect for human rights, inclusion, diversity, innovation and economic growth has been launched in June 2020.

## References

1. Ishimaru, S., et al.: Augmented learning on anticipating textbooks with eye tracking. In: Zlatkin-Troitschanskaia, O., Wittum, G., Dengel, A. (eds.) *Positive Learning in the Age of Information (PLATO) - A Blessing or A Curse?*, pp. 387–398. Springer, Wiesbaden, October 2017. [https://doi.org/10.1007/978-3-658-19567-0\\_23](https://doi.org/10.1007/978-3-658-19567-0_23)
2. Dengel, A.: Digital co-creation and augmented learning. In: *Proceedings KMO 2016, Hagen, 11th International Conference on Knowledge Management in Organizations, Hagen, Germany, July 2016*. ACM. 978-1-4503-4064-9/16/07. <https://dx.doi.org/10.1145/2925995.292605>
3. The Schumpeter Blog 2015 Homepage. <https://www.economist.com/news/business/21664190-modern-version-scientific-management-threatens-dehumanise-workplace-digital>. Accessed 01 June 2020
4. Devillers, L.: Social and emotional robots: useful artificial intelligence in the absence of consciousness. In: Nordlinger, B., Villani, C., Rus, D. (eds.) *Healthcare and Artificial Intelligence*, pp. 261–267. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-32161-1\\_32](https://doi.org/10.1007/978-3-030-32161-1_32)
5. Floridi, L., Cowls, J.: A unified framework of five principles for AI in society. *Harvard Data Sci. Rev.* **1**(1) (2019). <https://doi.org/10.1162/99608f92.8cd550d1>
6. Blandfort, P., Karayil, T., Borth, D., Dengel, A.: Captioning in the wild: how people caption images on Flickr. In: *ACM Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes (2017)*
7. Gomez, L., Mafla, A., Rusiñol, M., Karatzas, D.: Single shot scene text retrieval. In: *The European Conference on Computer Vision (ECCV)*, pp. 728–744 (2018)
8. Gómez, L., Rusiñol, M., Karatzas, D.: Cutting Sayre's Knot: reading scene text without segmentation. application to utility meters. In: *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 97–102. IEEE (2018)
9. Nayef, N., et al.: ICDAR2019 robust reading challenge on multi-lingual scene text detection and recognition—RRC-MLT-2019. In: *Proceedings 15th International Congress on Document Analysis and Recognition, IEEE CPS*, pp. 1582–1587 (2019)
10. Reddy, S., Mathew, M., Gomez, L., Rusiñol, M., Karatzas, D., Jawahar, C.V.: RoadText-1K: text detection recognition dataset for driving videos. In: *IEEE International Conference on Robotics and Automation (ICRA) (2020)*
11. Toyama, T., Dengel, A., Suzuki, W., Kise, K.: Wearable reading assist system: augmented reality document combining document retrieval and eye tracking. In: *Proceedings ICDAR 2013, 12th International Conference on Document Analysis and Recognition, Washington D.C., USA*, pp. 30–34, August 2013
12. Patel, Y., Gomez, L., Rusiñol, M., Karatzas, D.: Dynamic lexicon generation for natural scene images. In: Hua, G., Jégou, H. (eds.) *ECCV 2016. LNCS*, vol. 9913, pp. 395–410. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46604-0\\_29](https://doi.org/10.1007/978-3-319-46604-0_29)

13. Gomez, L., Patel, Y., Rusiñol, M., Jawahar, C.V., Karatzas, D.: Self-supervised learning of visual features through embedding images into text topic spaces. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
14. Gomez, R., Gomez, L., Gibert, J., Karatzas, D.: Self-supervised learning from web data for multimodal retrieval. In: Yang, M., Rosenhahn, B., Murino, V. (eds.) *Multimodal Scene Understanding: Algorithms, Applications and Deep Learning*, pp. 279–306. Elsevier (2019). ISBN 978-0-12-817358-9. <https://doi.org/10.1016/B978-0-12-817358-9.00015-9>
15. Mafla, A., Dey, S., Furkan Biten, A., Gomez, L., Karatzas, D.: Fine-grained image classification and retrieval by combining visual and locally pooled textual features. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2020)
16. Gomez, R., Gibert, J., Gomez, L., Karatzas, D.: Exploring hate speech detection in multimodal publications. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2020)
17. Lake, B.M., Ullman, T.D., Tenenbaum, J.B., Gershman, S.J.: Building machines that learn and think like people. *Behav. Brain Sci.* **40** (2017)
18. Terman, L.M.: *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*. Houghton Mifflin, Boston (1916)
19. Karayil, T., Irfan, A., Raua, F., Hees, J., Dengel, A.: Conditional GANs for image captioning with sentiment. In: Tetko, I., Kůrková, V., Karpov, P., Theis, F. (eds.) *Artificial Neural Networks and Machine Learning. Proceedings ICANN19, 28th International Conference on Artificial Neural Networks*, Munich, Germany, September 2019. LNCS, vol. 11730, pp. 300–312. Springer, Cham (2019). [https://doi.org/https://doi.org/10.1007/978-3-030-30490-4\\_25](https://doi.org/https://doi.org/10.1007/978-3-030-30490-4_25)
20. Narasimhan, M., Schwing, A.G.: Straight to the facts: learning knowledge base retrieval for factual visual question answering. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *ECCV 2018*. LNCS, vol. 11212, pp. 460–477. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-01237-3\\_28](https://doi.org/10.1007/978-3-030-01237-3_28)
21. Narasimhan, M., Lazebnik, S., Schwing, A.: Out of the box: reasoning with graph convolution nets for factual visual question answering. In: *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, pp. 2654–2665 (2018)
22. Hendricks, L.A., Burns, K., Saenko, K., Darrell, T., Rohrbach, A.: Women also snowboard: Overcoming bias in captioning models. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *ECCV 2018*. LNCS, vol. 11207, pp. 793–811. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-01219-9\\_47](https://doi.org/10.1007/978-3-030-01219-9_47)
23. Rohrbach, A., Hendricks, L.A., Burns, K., Darrell, T., Saenko, K.: Object hallucination in image captioning. In: *Empirical Methods in Natural Language Processing (EMNLP)* (2018)
24. Bhargava, S., Forsyth, D.: Exposing and correcting the gender bias in image captioning datasets and models. *arXiv preprint arXiv:1912.00578* (2019)
25. Biten, A.F., Gomez, L., Rusinol, M., Karatzas, D.: Good news, everyone! Context driven entity-aware captioning for news images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12466–12475 (2019)
26. Cheng, G., Ramirez-Amaro, G., Beetz, M., Kuniyoshi, Y.: Purposive learning: robot reasoning about the meanings of human activities. *Sci. Robot.* **4**, eaav1530 (2019)
27. Kuniyoshi, Y., Inaba, M., Inoue, H.: Learning by watching: extracting reusable task knowledge from visual observation of human performance. *IEEE Trans. Robot. Autom.* **10**, 799–822 (1994)
28. Ijspeert, A.J., Nakanishi, J., Hoffmann, H., Pastor, P., Schaal, S.: Dynamical movement primitives: learning attractor models for motor behaviors. *Neural Comput.* **25**, 328–373 (2013)

29. Bentivegna, D.C., Atkeson, C.G., Cheng, G.: Learning tasks from observation and practice. *IEEE Robot. Auton. Syst. J.* **47**, 163–169 (2004)
30. Ramirez-Amaro, K., Beetz, M., Cheng, G.: Transferring skills to humanoid robots by extracting semantic representations from observations of human activities. *Artif. Intell.* **247**, 95–118 (2017)
31. Tenorth, M., Beetz, M.: Representations for robot knowledge in the KnowRob framework. *Artif. Intell.* **247**, 151–169 (2017)
32. Heimerl, A., Baur, T., Lingenfesler, F., Wagner, J., André, E.: NOVA - a tool for eXplainable cooperative machine learning. In: *ACII 2019*, pp. 109–115 (2019)
33. Baur, T., Clausen, S., Heimerl, A., Lingenfesler, F., Lutz, W., André, E.: NOVA: a tool for explanatory multimodal behavior analysis and its application to psychotherapy. In: Ro, Y. M., Cheng, W.-H., Kim, J., Chu, W.-T., Cui, P., Choi, J.-W., Hu, M.-C., De Neve, W. (eds.) *MMM 2020. LNCS*, vol. 11962, pp. 577–588. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-37734-2\\_47](https://doi.org/10.1007/978-3-030-37734-2_47)
34. Zhao, R., Sinha, T., Black, A., Cassell, J.: Automatic recognition of conversational strategies in the service of a socially-aware dialog system. In: *Proceedings of the 17th Annual SIGdial Meeting on Discourse and Dialogue*, Los Angeles, CA, 13–15 September 2016 (2016)
35. Cassell: *The Ties that Bind: Social Interaction in Conversational Agents*. *Reseaux Issue number 220-221*, pp. 21-45 (2020)
36. Tartaro, A., Cassell, J., Ratz, C., Lira, J., Nanclares-Nogues, V.: Accessing peer social interaction: using authorable virtual peer technology as a component of a group social skills intervention program. *ACM Trans. Accessible Comput. (TACCESS)* **6**(1), 1–29 (2015). Article 2
37. Pecune, F., Chen, J., Matsuyama, Y., Cassell, J.: Field trial analysis of Socially Aware Robot Assistant. In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)* (2018)
38. Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., Cassell, J.: The effects of culturally-congruent educational technologies on student achievement. In: *Proceedings of Artificial Intelligence in Education (AIED)*, Memphis, TN, 09–13 July 2013 (2013)
39. Holstein, K., McLaren, B.M., Alevan, V.: Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *J. Learn. Anal.* **6**(2), 27–52 (2019). <https://doi.org/10.18608/jla.2019.62.3>
40. Holstein, K., McLaren, B.M., Alevan, V.: Designing for complementarity: teacher and student needs for orchestration support in ai-enhanced classrooms. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) *AIED 2019. LNCS (LNAI)*, vol. 11625, pp. 157–171. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-23204-7\\_14](https://doi.org/10.1007/978-3-030-23204-7_14)
41. Holstein, K., McLaren, B.M., Alevan, V.: Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In: Penstein Rosé, C., Martínez-Maldonado, R., Hoppe, H.U., Luckin, R., Mavrikis, M., Porayska-Pomsta, K., McLaren, B., du Boulay, B. (eds.) *AIED 2018. LNCS (LNAI)*, vol. 10947, pp. 154–168. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-93843-1\\_12](https://doi.org/10.1007/978-3-319-93843-1_12)
42. Ishimaru, S., Kunze, K., Kise, K., Dengel, A.: The wordometer 2.0: estimating the number of words you read in real life using commercial EOG glasses. In: *Proceedings UbiComp '16, 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, Germany, pp. 293–296. September 2016 <https://dx.doi.org/10.1145/2968219.2971398>
43. Tonomoto, S., Iwata, M., Kise, K.: Preliminary experiments toward personalized nudging strategies extensive reading of English. In: *Proceedings of CHI2020 Workshop on Detection and Design for Cognitive Biases in People and Computing Systems*, 8 p., April 2020



44. Ishimaru, T.M., Kise, K., Dengel, A.: Gaze-based self-confidence estimation on multiple-choice questions and its Feedback. Submitted to Asian CHI Symposium 2020, Honolulu, Hawaii, USA, April 2020. <https://doi.org/https://doi.org/10.1145/3391203.3391227>
45. Yamada, K., Augereau, O., Kise, K.: Estimation of confidence based on eye gaze: an application to multiple-choice questions. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (UbiComp 2017), pp. 217–220, 4 p. ACM, October 2017.
46. Grinbaum, A., Chatila, R., Devillers, L., Ganascia, J.G., Tessier, C., Dauchet, M.: Ethics in robotics research: CERNA mission and context. *IEEE Robot. Autom. Mag.* **24**(3), 139–145 (2017)
47. Devillers, L., et al.: Research ethics in Machine Learning, CERNA (2017). [https://www.al-listene.fr/files/2019/05/54730\\_cerna\\_2017\\_machine\\_learning.pdf](https://www.al-listene.fr/files/2019/05/54730_cerna_2017_machine_learning.pdf)