Chapter 9
Epilog

Toyoaki Nishida

Abstract Human-harmonized information technology is intended to establish basic technologies to achieve harmony between human beings and the information environment by integrating element technologies encompassing real-space communication, human interface, and media processing. It promotes a transdisciplinary approach featuring (1) the recognition and comprehension of human behaviors and real-space contexts by utilizing sensor networks and ubiquitous computing, (2) technologies for facilitating man–machine communication utilizing robots and ubiquitous networks, and (3) content technologies for analyzing, mining, integrating, and structuring multimedia data including those in text, voice, music, and images. It ranges from scientific research on the cognitive aspects of human-harmonized information processes to social implementations that may lead to breakthroughs in the harmonious interactions of human and information environments. In this chapter, I give an overview of achievements over the past eight years, remarking on insights as well as limitations. I also discuss future perspectives, singling out promising approaches.

Keywords Changing world · Computer and communication technology · Convivial society · Human and social potential · Human-harmonized information technology

9.1 A Brief History of Research on Human-Harmonized Information Technology

Our long and exciting journey was launched eight years ago by the late Professor Yoh'ichi Tohkura, based on a five-stage model, in which progress in the history of information in human society is conceptualized as consisting of the stages of

T. Nishida (✉)
Graduate School of Informatics, Kyoto University, Kyoto, Japan
e-mail: nishida@i.kyoto-u.ac.jp

T. Nishida
JST-CREST Research Area on Creation of Human-Harmonized Information Technology for Convivial Society, Tokyo, Japan

T. Nishida (ed.), *Human-Harmonized Information Technology, Volume 2*,
DOI 10.1007/978-4-431-56535-2_9

- the hunter–gatherer society, where information is used for creating and sharing knowledge about hunting to find food while avoiding dangers;
- the agricultural society, where information is used for sharing knowledge about growing and harvesting crops;
- the industrial society, where information is used for living to achieve material affluence;
- the information society, where information plays a central role in society, supporting our life; and
- the (prospective) convivial society, where information will be used to achieve harmony and enable creativity.

The present stands at the transition between the information society and the convivial society called the AI-supported society, in which people are surrounded by numerous autonomous intelligent agents that mediate between people and services.

AI-supported society may bring with it new problems, and an AI-supported society might eventually release human beings from the duties of supporting mankind on the planet. The transition phase from an information society to a convivial society may be painful. In the transition, people will need to find new areas of self-actualization. We will need to harmonize relationships between human beings and technology, not just by inventing human-friendly technology but also by making technology explore new forms of perception, activity, and creativity.

We consider human and social potential to be a central issue. Human potential is the power of an individual that enables her or him to actively sustain an endeavor to achieve a goal in maintaining a social relationship with other people. It involves vision, activity, sustainability, empathy, ethics, humor, and an esthetic sense. Vision permits one to initiate long-term coherent activity. Social potential is the power that a society of people possesses as a whole. It encompasses generosity, support, conviviality, diversity, connectedness, and innovativeness.

Human-harmonized information technology (IT) is intended for a convivial society, featuring perceptual information processing to harmonically interface between human beings and the information environment, featuring

- the recognition and comprehension of human behaviors and real-space contexts, utilizing sensor networks and ubiquitous computing;
- technologies for facilitating man–machine communication, utilizing robots and ubiquitous networks; and
- content technologies related to analyzing, mining, integrating, and structuring multi-media information processing.[1]

Supported by a research grant from the Japan Science and Technology Agency (JST),[2] seventeen research teams were selected, each of which was funded for slightly more than five years.

This work is now approaching its conclusion. Table 9.1 summarizes the current status of the projects. More than two-thirds have already been concluded, nine of

[1] http://www.jst.go.jp/kisoken/crest/en/research_area/ongoing/areah21-1.html.

[2] http://www.jst.go.jp/EN/index.html.

Table 9.1 Teams of the JST-CREST research area on Creation of Human-Harmonized Information Technology for convivial society

Duration	ID	Team
FY2009–2014	P1	*Life Log Infrastructure for Food (PI: Kiyoharu Aizawa)
	P2	*Dynamic Information Space based on High-speed Sensor Technology (PI: Masatoshi Ishikawa) ⇒ ACCEL Feasibility Study (2015) / ACCEL (FY2016–)
	P3	*Developing a communication environment by decoding and controlling implicit interpersonal information (PI: Makio Kashino)
	P4	*Smart seminar room based on multi-modal recognition of verbal and non-verbal information (PI: Tatsuya Kawahara)
	P5	*Elucidation of perceptual illusion and development of secse-centered human interface (PI: Yasuharu Koike)
	P6	*Sensing and controlling human gaze in daily living space for human-harmonized information environments (PI: Yoichi Sato)
	P7	*Modeling and detecting overtrust from behavior signals (PI: Kazuya Takeda)
	P8	*Construction and Utilization of Human-harmonized “Tangible” Information Environment (PI: Susumu Tachi) ⇒ ACCEL: Embodied Media Project (2014)
FY2010–2014	P9	*Studies on cellphone-type teleoperated androids transmitting human presence (PI: Hiroshi Ishiguro) ⇒ ERATO: Ishiguro Symbiotic Human-Robot Interaction Project (FY2014–)
FY2010–2015	P10	Development of a sound field sharing system for creating and exchanging music (PI: Shiro Ise)
	P11	Enabling a mobile social robot to adapt to a public space in a city (PI: Takayuki Kanda)
	P12	Development of Fundamental Technologies for Innovative Use of Character/Document Media and Their Application to Creating Human Harmonized Information Environment (PI: Koichi Kise)
	P13	Behavior Understanding based on Intention-Gait Model (PI: Yasushi Yagi)
FY2011–2016	P14	Building a Similarity-aware Information Environment for a Content-Symbiotic Society (PI: Masataka Goto) ⇒ ACCEL (FY2016–)
	P15	Pedagogical Machine: Developmental cognitive science approach to create teaching/teachable machines (PI: Kazuo Hiraki)
	P16	User Generated Dialogue Systems: uDialogue (PI: Keiichi Tokuda)
	P17	Harmonized Inter-Personal Display Based on Position and Direction Control (PI: Takeshi Naemura)

*: in the previous volume

which were reported in the previous volume and the remaining eight of which are presented in the preceding chapters in this volume. It is extraordinary that four teams have been upgraded, as one ERATO and three ACCEL projects supported by JST[3] have been launched, based on their achievements in the CREST research area.

In this chapter, I will summarize what has been achieved and examine its implications. I review the vision that I presented in our previous volume and discuss technical challenges and opportunities from a contemporary point of view.

9.2 What Has Been Achieved

Table 9.2 provides us with a bird's eye view of the entire contribution. This is an updated version of the one I presented in the previous volume. In this section, I will focus on achievements newly added since the previous volume.

9.2.1 Platform Research Level

The platform research level is a technical core of the human-harmonized IT. Our focus was interaction at the perceptual level.

In the previous volume, Tachi and his colleagues presented a rather comprehensive suite of platforms for a tangible information environment [21], including a number of haptic information displays, an introductory haptic toolkit, a vision-based thermal sensor, a telexistence avatar robot system that can provide the experience of an extended body schema, a retro-reflective projection technology (RPT)-based full-parallax autostereoscopic 3D that can generate vertical and horizontal motion parallax, and an autostereoscopic display that can produce a 3D image in mid-air with a view of 180 degrees.

Ishikawa and his colleagues presented four technologies: high-speed 3D vision for insensible dynamics sensing, a high-speed resistor network proximity sensor array for detecting nearby objects, non-contact low-latency haptic feedback, and high-speed display of visual information toward creating a dynamic information space that could harmonize the human perception system, recognition system, and motor system [9]. They also demonstrated how a system of a dynamic information space could be developed by using these four technologies. Their demonstration included an AR typing interface for mobile devices and a high-speed gaze controller for high-speed computer-human interactions, as well as more integrated systems, such as VibroTracker, a vibrotactile sensor for tracking objects, and AIRR Tablet, a floating display with a high-speed gesture user interface.

Sato and his colleagues described a suite of techniques to sense and control human gazes not only in a laboratory environment but also in a living space [19]. They intro-

[3] https://www.jst.go.jp/kisoken/accel/en/index.html and https://www.jst.go.jp/erato/en/index.html.

Table 9.2 A bird's eye view of the Human-Harmonized Information Technology in this volume

Category and features
Social implementation
• *Field studies on *sonzaikan* media: eldery care and education support [8] • *Field studies with FoodLog Web and app [1] • *Public installation at Miraikan • Field trials of robot services at public space [10] • Public installation at Nagoya Institute of Technology and Handa City [23]
Application
• *A suite of *sonzaikan* media: Telenoid, Hugvie, and Elfoid [8] • music-sync applications using Songle Widget, TextAlive and many other web services [5] • SHelective, Inter-Personal Browsing, PVLC projector, EnchanTable, MiragePrinter, and fVisiOn [16] • *FoodLog Web with image processing for food-balance estimation and FoodLog app for assisting food recording by image retrieval [1]
Platform
• *Tangible information environment encompassing haptic information display, Telexistence avatar robot system: TELESAR V, RPT-based full-parallax autostereoscopic 3D: RePro3D, and an autostereoscopic display for seamless interaction with the real environment mixed together: HaptoMIRAGE [21] • *Dynamic information space based on integrated high-speed 3D vision for insensible dynamics sensing, high-speed resistor network proximity sensor array for detecting nearby object, noncontact low-latency haptic feedback, and high-speed display of visual information [9] • Intention-Gait model [26] • Human-robot symbiosis technology based on high precision pedestrian sensing and modeling [10] • *Technology for sensing and controlling human gaze in daily living space [19] • Reading-life log [13] • Songle and Songle Widget, geared by musical similarity and typicality estimation technologies [5] • MMDAgent [23] • Sound cask [7] • *Smart Posterboard for archiving poster conversation [12]
Basic science
• Finding of sensitivity to temporal contiguity in pedagogical interaction [6] • *Cognitive model of excessive trust [22] • **Sonzai* and *sonzaikan*: cognitive model of sense of presence [8] • *Implicit interpersonal information in communication [11] • *Haptic primary color model [21] • *Conceptual model of sense-centered human interface based on illusions in perceived heaviness [14] • *Multi-modal corpus for poster conversation [12]

*: in the previous volume

duced and implemented an appearance-based gaze-sensing method with adaptive linear regression (ALR), which can make an optimal selection of a sparse set of training samples for gaze estimation, a new approach to the auto-calibration of gaze sensing from a user's natural viewing behavior, which was predicted with a computational model of visual saliency, and (3) user-independent single-shot gaze estimation. They also studied a subtle modulation of visual stimuli based on visual saliency models and the use of a robot's nonverbal behaviors in human–robot interaction.

Kawahara and his colleagues gave useful insights into the prediction of turn-taking, speaker dialization, hot-spot detection, and the prediction of interest and comprehension levels by analyzing multi-modal conversations using the Smart Posterboard they developed [12]. For example, it was found that about 70% of next speakers in turn-taking events could be predicted by combining eye-gaze objects, joint eye-gaze events, duration, and backchannels. It was also found that interest levels could be predicted by using the occurrence of questions and prominent tokens and comprehension levels could be estimated from question types.

In this volume, Yagi and his colleagues focused on the varieties of gait a person has to determine the relation between gait variation and inertial states, i.e., attention (gaze direction), human relations (group segmentation), and cognitive level (assessment of dementia) [26]. For attention estimation, they conducted numerous experiments studying the relationship between gaze and whole-body behaviors. They found that there is a similar eye–head coordination in different conditions, which suggests that head orientation is directly related to visual perception; the distribution of the eye position varies systematically with head orientation; the angles of the gaze, head, and chest have linear relationships, under non-walking and walking conditions; not only the head but also the arm and leg movements are related to the gaze locations; etc. They also propose a method for determining whether two people belong to the same group, combining motion trajectory, chest orientation, and gesture. Researching dual-task analysis for cognitive level estimation, they conducted data collection at an elderly-care facility and at the National Museum of Emerging Science and Innovation, or Miraikan, in Tokyo. The data obtained from the latter are immense, with more than 95,000 participants. The analysis of these data is in progress.

Kanda and his colleagues presented a project intended to enable a mobile social robot to adapt to an open public space in a city [10]. To harmonize mobile social robots in daily human contexts, they address common-sense problems in the domain of open public spaces, such as shopping mall corridors, where pedestrians do not just walk but stop to interact, stop to observe, slow down to look, and leave uninterested. They built a comprehensive pedestrian model for human–robot symbiosis. This includes collision avoidance and task-oriented HRI, encompassing such activities as shopping and observation. It can predict the behavior of people in low-density situations (less than 0.25 person/m^2). They also introduced high-level harmonized HRI features to avoid collision, prevent congestion, and escape "robot abuse"—the nasty treatment of robots, by children in particular. They conducted several field studies and found that they were able to harmonize mobile robots in daily human contexts and encouraged people to acquire information from them.

Kise and his colleagues presented reading-life log technology to help people live with characters [13]. Their technology is comprehensive, including real-time character recognition for alpha-numeric and Japanese characters, omnidirectional character recognition that allows the recognition of all characters in a 360-degree scene image, and real-time document-image retrieval based on basic character detectors and recognizers, a large-scale character dataset, and an automatic font generator. The reading-life log allows not only the construction of a record of one's reading life, including the time spent on, amount of, and attitude toward reading activities, but also the analysis of the content of the reading to support the user's intellectual activities. Using reading-life log technology, they prototyped applications such as Wordometer, which counts the number of words one reads to diagnose one's reading life, a scene-text detector and generator, an automated text annotator, a system for recording texts together with the facial expressions of the reader, and an augmented narrative that uses bio-feedback in a text–body interaction.

Goto and his colleagues presented a suite of technologies for building a similarity-aware information environment [5]. Songle is a web-based active music appreciation service that can automatically determine four types of musical descriptions: musical structure (chorus sections and repeated sections), hierarchical beat structure (musical beats and bar lines), melody line (fundamental frequency, f0 of the vocal melody), and chords (root note and chord type). Songle's web service permits anonymous users to correct errors in the musical archive, to cope with the incompleteness of the automated tool. Songle Widget is a web-based multimedia development framework that allows for the control of computer-graphic animation and physical devices such as robots in synchronization with music publicly available on the web.

Tokuda and his colleagues describe a framework for user-generated content creation [23]. They developed the MMDAgent toolkit to build speech-interaction systems for consumer-generated media (CGM). This allows the building of voice-interaction systems incorporating speech recognition, HMM-based flexible speech synthesis, embodied 3D agent rendering with simulated physics, and dialogue management based on a finite-state transducer. MMDAgent was released as an open-source software toolkit. Tokuda and colleagues have constructed the all-in-one Encyclopedia MMDAgent package, integrating a set of materials on the use of MMDAgent and the production of dialogue content, including guide books/tutorials, slides, reference manuals, and sample scripts.

Ise and his colleagues addressed 3D-sound-scene reproduction [7]. They have succeeded in the world's first implementation of an immersive auditory display, named the Sound Cask, which implements the principle of boundary surface control (BoSC), a theory of 3D-sound-field reproduction. BoSC features the ability to reproduce a sound field, not by points but in three dimensions. Sound Cask was designed to be a wooden output device based on BoSC theory, considering the reflection from the wooden frame, strong normal modes of the outer rectangular chamber, and the wooden distribution of loudspeakers. The system can provide high performance of spatial information reproduction, including sound localization and sound distance, even as the listener freely moves her or his head. A performance evaluation of the system is reported, which encompasses physical performance, localization, and the psychological and physiological evaluation of the feeling of reality in a 3D sound field. The Sound Cask system helps music professionals such as musicians, acoustic engineers, music educators, and music critics enhance their skills and further explore their creativity by providing them with the means to experience 3D sound in a telecommunications environment. Applications include a sound-field simulator, sound table tennis, and sound-field sharing. Sound Cask has been evaluated from various angles: physical performance, localization, and psychological and physiological aspects of feeling reality in a 3D field.

9.2.2 Basic Science Level

The basic science level of the human-harmonized IT addresses basic scientific issues for supporting the platform level. It involves human perception that underlies inter-human and human-agent interactions.

In the previous volume, Takeda and his colleagues presented a cognitive model of excessive trust in human cognition and behaviors [22]. They demonstrated that an integrated model of the gaze and vehicle operational behavior was effective for detecting risky lane changes. Ishiguro and his colleagues introduced *sonzaikan*[4] media, combining auditory and tactile sensations to represent the feeling of presence in the challenge to transmit human presence [8]. A minimal design approach was employed so that the user can project her or his own image of the communication partner on the terminal device. Kashino and his colleagues employed the idea of implicit interpersonal information (IIPI) to decode mental states, identify the cause of impaired communication in high-functioning autism spectrum disorder (ASD), improve the quality of communication, and elucidate neural mechanisms involved in the processing of IIPI [11]. Tachi and his colleagues proposed a haptic primary-color model to serve as the foundation for designing a haptic information display to recreate cutaneous sensation [21]. Their study draws on studies of how information is mapped between physical, physiological, and psychological spaces. Koike and his colleagues studied the illusion of perceived heaviness induced by the time offset between visual and haptic contact [14]. Their interesting findings included that an object was perceived to be heavier when force was applied earlier than visual contact and perceived to be lighter when it was applied later. They introduced the point of subjective simultaneity (PSS) and the point of subjective equality (PSE) to quantitatively measure the subjective evaluation of timing and weight perception.

In this volume, Hiraki and his colleagues addressed pedagogical machines that can teach and be taught, as teaching and learning behaviors are central to human intelligence. They took a three-fold approach: the development of cognitive science, machine intelligence, and field studies in an educational environment. In the first approach, they found that infants are very sensitive to temporal contiguity in interac-

[4]The sense of presence. It is present when its presence is recognized by a person. In contrast, an accompanying concept, *sonzai* refers to an objective presence.

tion with their mothers. In particular, it became clear that nowness and responsiveness are very important for the design of a pedagogical machine. They investigated temporal contiguity in pedagogical situations and found from a controlled experiment that a 1-second delay in presenting an image significantly deteriorated imitation scores. With the second approach, they developed a pedagogical agent with gaze interaction (PAGI) that is designed to teach Korean words to Japanese students, capable of simulating mutual gaze, gaze following, and joint attention. Experiments with PAGI revealed that even adults are implicitly affected by nowness and responsiveness during word learning with artificial agents. As a result of the third approach, several novel findings have been obtained, e.g., several interactions among children where the interactions seem to affect each child's learning and altruistic behaviors.

9.2.2.1 Application Level

In contrast to the basic science level, the application level of the human-harmonized IT is oriented toward building engineered artifacts to benefit society.

In the previous volume, Ishiguro and his colleagues presented a suite of *sonzaikan* media, consisting of Telenoid, Hugvie, and Elfoid. Telenoid was created as a test-bed based on the minimal design of a human being. Hugvie is a human-shaped cushion phone. Human-likeness in visual and tactile information was emphasized in Telenoid's design to facilitate human–robot and mediated inter-human interactions. Hugvie focused on human voice and a human-like touch. Elfoid is a hand-held version of Telenoid. The cellular phone version can connect to a public cellular phone network and was designed to provoke stronger *sonzaikan* than normal cellular phones. The underlying technologies include motion generation through speech information and motion generation and emotional expression through visual stimuli.

Aizawa and his colleagues described FoodLog Web and an app. FoodLog Web is a system that not only allows the user to create a food log by simply taking a photograph of what they have eaten but also applies image processing to analyze the uploaded photograph to generate food-balance information to enable food assessment [1]. The FoodLog app runs on smart phones, allowing the use of photographs to easily add textual descriptions.

In this volume, Goto and his colleagues listed various music-sync applications using Songle Widget, which has been made open to the public [5]. They include a two-dimensional music-sync animation, three-dimensional animated dancing characters, music-sync lighting, a video-jockey web service, real-time control of music-syncro bot dancing, a music visualizer featuring three-dimensional music-sync animation, and a music-sync photo slideshow. Songrium is a music-browsing assistance service that allows the visualization and exploration of a large amount of user-generated music content. They also developed content-creation support technologies. TextAlive enables the creation of music-synchronized lyrics animation. This makes it easy to produce lyric animations based on online music and lyrics information. It supports creativity with intuitive interfaces and mechanisms that simplify the production of derivative work, and it enables the enhancement of effects and incorporation of editing features by web-based programming. It should be noted that Goto and colleagues have developed numerous other applications.

Naemura and his colleagues addressed the privacy control of display content to promote discussion in groups, projection-based control of physical objects for suppressing the incompatibility between the physical and digital worlds, and spatial imaging for augmented reality among people without wearable displays [16]. For privacy-control issues, they developed a privacy-control method called SHelective for sharing displays, and a group-work facilitation system called Inter-Personal Browsing for collaborative web search. For the projection-based augmentation issue, they propose the concept of a bit-data projection system called the Pixel-level Visible Light Communication (PVLC) Projector, and a chemical augmentation system called Hand-Rewriting for paper-based computing. These latter two are functional extensions of existing image projectors to realize a more advanced augmentation of the physical world. For spatial imaging, they propose EnchanTable, which can display a vertically standing mid-air image on a table surface using reflection, MiragePrinter for interactive fabrication on a 3D printer with a mid-air display, and fVisiOn for a glasses-free tabletop 3D display viewable from 360 degrees to augment ordinary tabletop communications.

9.2.3 Social Implementation and Field Study

The social implementation level of the human-harmonized IT addresses implementation of technology as a part of a social system. It includes public installations and field studies.

In the previous volume, we presented four long-term exhibitions at the National Museum of Emerging Science and Innovation (Miraikan)[5] by Naemura, Ishiguro, Tachi, Yagi, and their colleagues as a valuable outreach to society. The first exhibition, by Naemura and his colleagues, entitled "The Studio—Extend Your Real World—"[6] demonstrated display design in a mixed-reality environment, mounted for 164 days from July 3, 2013, to January 13, 2014. It attracted about 130,000 visitors. The second exhibition, by Ishiguro and his colleagues, has been sustained for over two and half years since June 25, 2014. The exhibition, entitled "Android: What is Human?"[7] demonstrating interaction with androids and a Telenoid, attracted enormous public attention. The estimated number of visitors before March 2014 was more than 500,000. The third exhibition, by Tachi and his colleagues, was open under the title "Touch the World, Feel the Future," from October 22, 2014, to May 11, 2015.[8] The installation includes a telexistence robot TELESAR V, a 3D-display HaptoMIRAGE, which allows the user to touch something that they see exactly as

[5] http://www.miraikan.jst.go.jp/en/.

[6] http://miraikan.jp/medialab/en/12.html.

[7] https://www.miraikan.jst.go.jp/en/exhibition/future/robot/android.html.

[8] http://miraikan.jp/medialab/en/14.html.

they see it, Haptic Broadcast, which can transfer tactile sensations through the activation of sport apparatus in synchronization with audio–visual information, Haptic Search, which allows searching for an object with the sense of touch similar to a given stimulus, the TECHTILE toolkit, which allows a user to design his or her own senses of touch easily, and GravityGrabber, which can produce normal and tangential forces on a fingertip. Around 140,000 people visited in the first half of the exhibition period. The fourth, launched by Yagi and his colleagues on July 15, 2015,[9] was about behavior understanding, based on an intention-gait model. The exhibition, originally planned to go until April 11, 2016, was extended to June 27. More than 95,000 people participated in the experiment in the installation.

We described field studies on *sonzaikan* media conducted by Ishiguro and his colleagues. Elder care with Telenoid, cultural differences toward Telenoid, and educational support with Telenoid were investigated in field studies. We also described the FoodLog web service[10] that Aizawa and his colleagues developed, which not only allows the user to create a food log simply by taking a photo of what she or he eats but also generates food-balance information through image analysis [1].

In this volume, the list of longitudinal public installations is extended. Numerous web services have been made available, as reported by Goto and his colleagues [5]. Tokuda and his colleagues demonstrated their uDialogue technology through a public installation. They developed a bidirectional voice-guidance digital signage system ("Mei-chan") using the MMDAgent toolkit and installed it in an open space just outside the main gate of Nagoya Institute of Technology. They also built an indoor digital signage system capable of bidirectional voice guidance, installing it at a tourist information office for public service in Handa city, Aichi, Japan [23]. Kanda conducted two field trials evaluating robot service in a public space. In the flyer distribution trial, they developed a detailed behavioral model of people in the flyer-distribution service, as well as analyzing successful behaviors. Thanks to these efforts, the success ratio of the robot distributor reached 87.9%, exceeding that of average human distributors (40.7%). In the direction-giving trial, 96.6% of visitors reported that the robot's service was good; people tended to listen to the robot and seemed happy during the interaction [10].

9.3 Present Situation

In the year since the previous volume, weak AI, or AI as an intelligent tool, has become even more popular in our living spaces, as seen in talking assistants, chatbots, and semi-autonomous vehicles using autopilot, in addition to epoch-making events, such as AlphaGo's win against Lee Sedol.

Strong AI, or AI as a conscious autonomous agent, is now visible on the horizon, though currently beyond implementation. Although strong AI appears solely in

[9] http://www.miraikan.jst.go.jp/en/exhibition/future/digital/medialabo.html.

[10] http://www.foodlog.jp/en.

science fiction, as an extrapolation of existing weak AIs, people perceive it rather concretely, either as it is written about in stories or depicted in movies.

The mixture of weak and strong AI has provoked social concerns. Some of these concerns are pressing, such as safety and controllability, for example in autonomous vehicles, for their behaviors are not completely predictable or continuous as a function of situations. Other concerns are imaginary but not innocent, as they provoke anxiety in society.

A pair of open letters from the Future of Life Institute (FLI) concisely represents social concern with the current status of artificial intelligence. The first open letter [4] openly proposes that AI research should not only focus on technological advances in artificial intelligence but also make maximal effort to ensure AI is beneficial to society. In the detailed version of a report published in AI Magazine [18], Russell et al. present a rather comprehensive list of research priorities for socializing artificial intelligence. They emphasize verification, validity, security, and control as computer-science research aspects of AI. The second open letter [3] claims that autonomous weapons using AI technology should be prevented for the sake of humanity. AI is even counted among the twelve risks that threaten human civilization [17].

Long-term issues have been discussed in different contexts. The report of the 2015 study panel on a 100-year study on artificial intelligence [20] takes the positive approach that there is "no cause for concern that AI is an imminent threat to humankind," whereas AI technologies have the "potential to be used for good or nefarious purposes." This motivates us to steer AI research and development so that it benefits society.

Machine ethics is the study of giving machines ethical principles or procedures for autonomously judging whether a given action is good [2]. So, long as machines are living with us, the more autonomy the machines have, the more ethical they need to be. Moor classifies ethical agents into four categories, namely normative agents that are not designed for ethical situations, ethical impact agents whose behavior may have an ethical impact regardless of their task, implicit ethical agents that have been programmed to either support ethical behaviors or avoid unethical behaviors, and explicit ethical agents that can calculate the best action in ethical dilemmas [15].

The influence of technology on mankind may be more implicit, quietly going on beneath the surface and hard to observe without insight and professional investigation. In her best-selling book Alone Together [25], Turkle analyzes digital culture from a psychoanalytic viewpoint. She focuses on robotic creatures and networked life. On the one hand, Turkle points out, digital creatures invoke a sense of companionship even though they do not understand the user. This may be regarded as moral complacency resulting from a non-authentic relationship with an artificial caregiver; as she wrote, "robots cannot pretend because it can only pretend."[11] With regard to other people, Turkle discusses the multi-tasking brought about by the digital network that may lead to getting "alone together" even if they physically live together. Technology always brings about negative effects as well as positive ones. Sometimes, the

[11] p. 124 in [25].

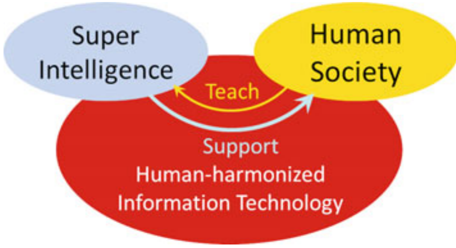


Fig. 9.1 Role of human-harmonized information technology

former silently spreads beneath the surface. People are vulnerable to loneliness and hence to artificial togetherness.

We are convinced that movement toward human-harmonized IT is well-supported and on the right track. We have to date only focused on the perception level. For the next step, we will need to shed light on the higher level and aim for integrated services. In addition to the treatment of semantic and contextual information, we will need to address the huge amount of information on the net. Goto has made an initial attempt in this direction in the Songrium project. Generalization is needed.

9.4 Potential Next Challenges

What are the next challenges? As we witness the rise of superintelligence accelerating the shift from an information society to an AI society, more people are coming to share the view that human society will benefit from superintelligence, through learning (Fig. 9.1). The role of human-harmonized IT is realizing the intensive integration between human society and super intelligence, so that human and social potential may be maximized. In this chapter, we will draw a clear picture of integration and identify technical challenges and opportunities to make the maximization happen.

9.4.1 Companion Agent

Among other potential contributions to human-harmonized IT, the most integrated image of the target may be the companion agent (Fig. 9.2), which works with each individual user by exploiting functions provided by the social platform geared by superintelligence. The role of a companion agent is to provide a personalized comprehensive support for the user, not just by protecting her or him from potential threats in the jungle of complex functions available from numerous services built on the social platform, but also by enhancing her or his human potential to maximally leverage and contribute to social potential.

The notion of companionship is much more subtle and sophisticated than assistance. Unlike an intelligent assistant, a companion agent involves numerous con-

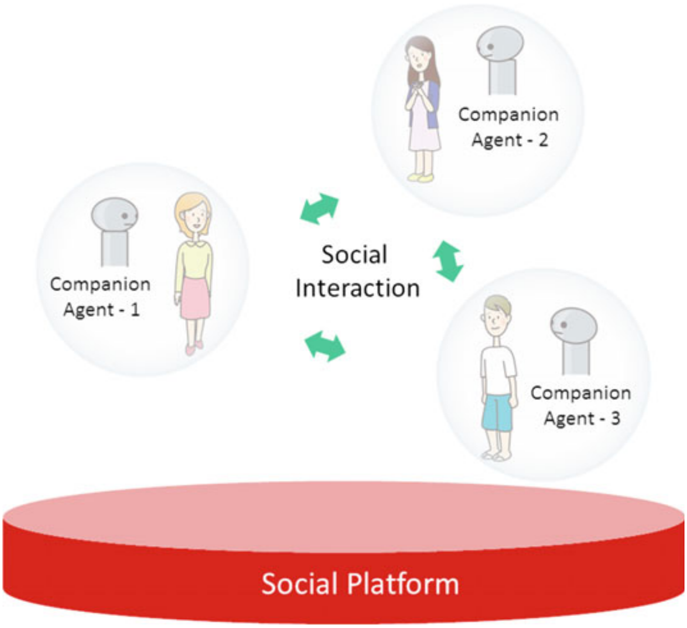


Fig. 9.2 Companion agents

ceptual challenges as well as technical ones. David Traum [24] suggested that a companion may be characterized by a long-term relationship such as unquestionable loyalty, second banana/sidekick, and interest in its prime or boss. Traum suggests that companions should be carefully distinguished from similar concepts such as servants, agents, leaders, and counselors as well as assistants and even romantic partners. Traum lists ten dos and don'ts for artificial companionship:

1. Do something useful!
2. Don't just do one thing!
3. Do understand the situation (as much as possible)!
4. Do what I want!
5. Be a good conversationalist!
6. Do revisit conversations appropriately!
7. Do understand what you are saying!
8. Don't be repetitive!
9. Do learn and customize!
10. Do maintain immersion!

Furthermore, ethical issues need to be addressed, even though companions are not explicitly at the ethical-agent level but implicitly at the ethical-agent level [15]. Ethical issues come in at various stages in scenes.

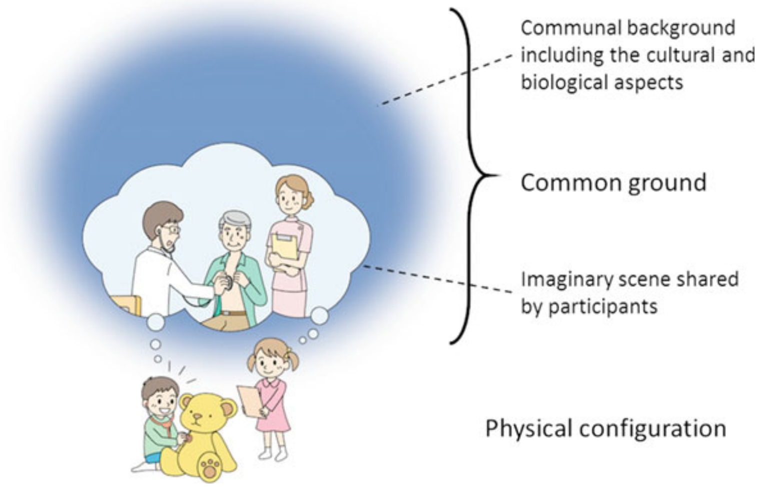


Fig. 9.3 Common ground

9.4.2 Common Ground

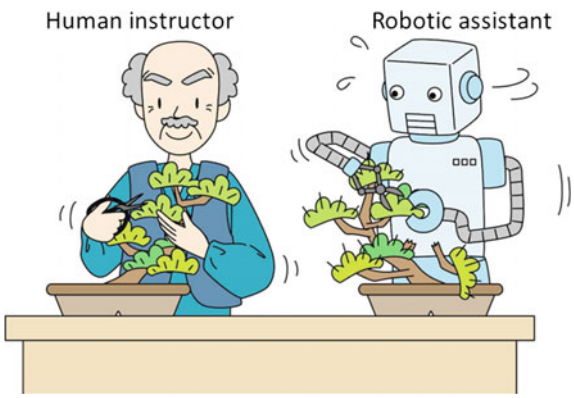
Common ground is the collection of knowledge, beliefs, and suppositions that are required that each participant in an interaction share prior to interaction. It ranges from communal background, including a cultural and biological basis, to an imaginary scene dynamically formed, shared, and updated as interaction proceeds (Fig. 9.3). Unless extensive common ground is shared, interactants must spend immense effort coordinating their actions and sharing ideas.

Common ground ranges from the perceptual to the story levels. The implicit nature of common ground makes its computation difficult. Even if immense amounts of data are available on the net and sensory data are measurable by advanced sensors, these are useless unless they are organized and associated with the given situation.

9.4.3 Robotic Apprentice

Building a robotic apprentice that can mimic a human instructor may be a promising approach to acquiring common ground regarding actions and interactions (Fig. 9.4). Humanoid robots may be employed to transfer human knowledge and expertise rooted in embodiment, ranging from daily activities to highly skilled behaviors. Humanoid robots with human-like perceptions are mandated for this task, as the apprentice needs to be able to capture the world and its tasks in a very similar fashion to human instructors. Put another way, robotic apprentices for such tasks

Fig. 9.4 Robotic apprentice



must possess common ground at a perceptual and motor level in addition to basic language-communication capabilities.

Robots with different levels of embodiment may be used in cases where we need to transfer task-related knowledge at a more abstract level. This setting is more challenging, as the robot must be able to transform the perceptions and behaviors of the user to a different embodiment of the robot.

In addition, a powerful machine-learning algorithm is necessary to generalize the robotic apprentice's experience into a more general form of knowledge for incorporation into a larger body of knowledge for reuse.

9.4.4 From Perception to Cognition

Both common-ground building and robotic apprentices require multiple levels of information to be integrated, because these tasks need a full range of communication with people. Two challenges arise. The first is addressing modalities other than speech and vision. People interact with the environment and each other not just through audio-visual channels but also through haptic, tangible, smell, taste, temperature, and other sensations. People use a rich vocabulary regarding these sensations, sometimes single and sometimes composite, to communicate their experiences. The second challenge is to create a solid link between perception and cognition where high-level issues such as understanding, storytelling, and memory organization matter. I believe that this will also benefit the machine ethics that should accompany practice in daily life.

9.4.5 Helping People Grasp Huge Information Space

A difficulty for the information society and its successors is the explosion of information, which makes it desperate for people to gain the sense of the "whole" space, which may lead people to ignorance or anxiety and significantly hinder creativity and other forms of human potential. An important challenge to human-harmonized IT is helping people gain the sense of the "whole" in the huge information space, so they are able to feel its benefits and risks. A less ambitious but still useful challenge is to produce a historical development of an idea in the information space, e.g., the history of artificial intelligence research, on demand. As human beings cannot spend an infinite amount of time answering a question, the ability to address such historical development questions would be a valuable function for a companion agent.

9.4.6 Gamification and Storytelling

In the convivial society, creativity will become a central concern of life. Gamification and storytelling are keys to creativity. Our companion agent must be able to help people gamify and tell stories. Both games and stories are great inventions of the humanity and are lots of fun to people. They are complementary to each other, shedding light on different aspects of reality. Games abstract the reality as interactions with rules, while stories allow people to capture it as relations. We can characterize the relationship between the reality, games and stories as in Fig. 9.5. Gamification is a process of abstracting the real world by eliciting the rules for competition, whereas storytelling is a process of organizing experiences, either real or gamified, into a structure. Conversely, the science of the artificial allows us to pave the world as we think in stories and games.

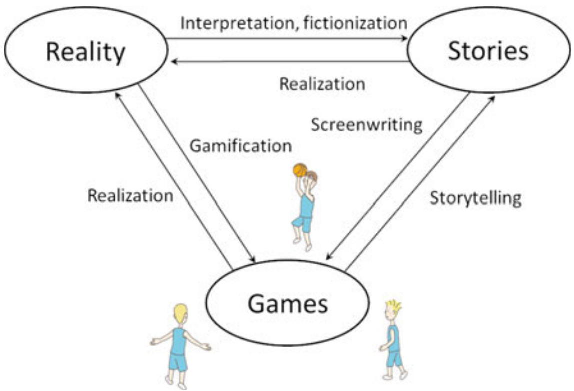


Fig. 9.5 Gamification and Storytelling

9.4.7 Longitudinal Companionship

The sharp difference between companionship and mere assistance is that companionship must be longitudinal by definition. For a valuable companion agent, one-time assistance is less critical than long-term sustaining support and underlying trust and emotional attachment. Even if a companion agent makes a mistake or an error, it can be repaired. However, social and emotional relationships matter. Any negative actions that may hurt a social or emotional relationship, such as betrayal or insults, may immediately lead to the breakdown of companionship. The challenge is how to build a long-term mutual trust, incorporating repair and resilience functions. Companionship must also be designed to retain human potential that may even lead to the break-up of companionship. Ethics must cope with such negative occurrences.

9.4.8 Practice and Social Implementation

It is always important for us to learn from people through practice. There are several issues for us in designing our research for social implementation. Apart from specialized subjects such as medical clinics, scientific research, and athletics, we need to choose subject domains carefully. We believe that daily-life support is not only demanded but also allows us to learn a great deal, though it is very challenging in the sense that the boundary is hard to find and information is unstructured and tacit.

Methods of evaluation and assessment are crucial for learning from feedback. In addition to standard means, such as brochures or physiological indices, it is instructive to explore new approaches, as, for instance, it was found that the measurement of neuroendocrine responses is useful for probing the effects of social interactions [8]. Further investigation is expected to greatly contribute to the evaluation and assessment of companion agents.

9.5 Conclusion

This chapter gave an overview to the JST-CREST research area on the creation of human-harmonized IT for a convivial society. I emphasized changing-world phenomena as a background and characterized our research area as a challenge for developing key technology for a convivial society in which humans and technology are in harmony. I referred to the late Professor Tohkura's grand conjecture concerning the five-stage shift in the role of information in society, and pointed out that we are at the stage of an information society and are moving toward an AI-supported society. I argued that a key idea in making a successful transition to a convivial society is through human and social potential. Human potential is the power of an individual that enables her or him to actively sustain an endeavor to achieve a goal in maintaining a social relationship with other people. It involves vision, activity, sustainability, empathy, ethics, humor, and esthetic sense. Social potential is the power that a society of people possesses as a whole. It encompasses generosity, support, conviviality, diversity, connectedness, and innovativeness. We believe that our research area contributes to the construction of technology that will enhance human and social potential. The outcomes from the first group encompass a suite of topics ranging from foundation to social implementation, covering novel subjects such as implicit interpersonal information, sense-centered human interfaces, excessive trust, and sense of presence (*sonzaikan*). Applications include FoodLog, which is a suite of *sonzaikan* media that has been socially implemented through field trials.

References

1. K. Aizawa, FoodLog: multimedia food recording tools for diverse applications (2016). (in volume 1)
2. M. Anderson, S.L. Anderson (eds.), *Machine Ethics* (Cambridge University Press, Cambridge, 2011)
3. Future of life institute. Autonomous weapons: an open letter from AI and robotics researchers (2015), http://futureoflife.org/open-letter-autonomous-weapons/
4. Future of life institute. Research priorities for robust and beneficial artificial intelligence (2015), http://futureoflife.org/ai-open-letter/
5. M. Goto, Building a similarity-aware information environment for a content-symbiotic society (2016). (in this volume)
6. K. Hiraki, Pedagogical machine: Developmental cognitive science approach to create teaching/teachable machines (2016). (in this volume)
7. S. Ise, Deelopment of a sound field sharing system for creating and exchanging music (2016). (in this volume)
8. H. Ishiguro, Transmitting human presence through portable teleoperated androids—a minimal design approach (2016). (in volume 1)
9. M. Ishikawa, I. Ishii, Y. Sakaguchi, M. Shimojo, H. Shinoda, H. Yamamoto, T. Komuro, H. Oku, Y. Nakajima, Y. Watanabe, Dynamic information space based on high-speed sensor technology (2016). (in volume 1)
10. T. Kanda, Enabling a mobile social robot to adapt to a public space in a city (2016). (in this volume)
11. M. Kashino, S. Shimojo, K. Watanabe, Critical roles of implicit interpersonal information in communication (2016). (in volume 1)
12. T. Kawahara, Smart posterboard: multi-modal sensing and analysis of poster conversations (2016). (in volume 1)
13. K. Kise, Reading-life log as a new paradigm of utilizing character and document media (2016). (in this volume)
14. Y. Koike, Elucidation of perceptual illusion and development of sense-centered human interface (2016). (in Volume 1)
15. J.H. Moor, The nature, importance, and difficulty of machine ethics, in *Machine Ethics*, ed. by M. Anderson, S.L. Anderson (Cambridge University Press, New York, 2011), pp. 13–20. Cambridge Books Online
16. T. Naemura, Inter-personal displays: augmenting the physical world where people get together (2016). (in this volume)
17. D. Pamlin, S. Armstrong, Global challenges: 12 risks that threaten human civilization (Global Challenges Foundation, Stockholm, 2015)

18. S. Russell, D. Dewey, M. Tegmark, Research priorities for robust and beneficial artificial intelligence. AI Mag. **36**(4), 105–114 (2014)
19. Y. Sato, Y. Sugano, A. Sugimoto, Y. Kuno, H. Koike, Sensing and controlling human gaze in daily living space for human-harmonized information environments (2016). (in volume 1)
20. P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyanakrishnan, E. Kamar, S. Kraus, K. Leyton-Brown, D. Parkes, W, Press, A. Saxenian, J. Shah, M. Tambe, A. Teller, Artificial intelligence and life in 2030, 2016. One hundred year study on artificial intelligence: report of the 2015-2016 Study Panel (Stanford University, Stanford, CA, September 2016), http://ai100.stanford.edu/2016-report. Accessed 1 Aug 2016
21. S. Tachi, Haptic Media: construction and utilization of haptic virtual reality and haptic telexistence (2016). (in volume 1)
22. K. Takeda, Modeling and detecting exceesive trust from behavior signals: Overview of research project and results (2016). (in volume 1)
23. K. Tokuda, User generated dialogue systems: uDialogue (2016). (in this volume)
24. D.R. Traum, Do's and don'ts for software companions. In *Proceedings of the International Workshop on Emotion Representations and Modelling for Companion Technologies*, ERM4CT '15 (ACM, New York, NY, USA, 2015), pp. 3–3
25. S. Turkle, *Alone Together: Why we expect more from technology and less from each other* (Basic Books, New York, 2011)
26. Y. Yagi, Behavior understanding based on intention-gait model (2016). (in this volume)