ISSN 2186-7437

NII Shonan Meeting Report

No. 188

Intelligent Interaction with Autonomous Assistants in the Wild

Yutaka Arakawa
Wolfgang Minker
Elisabeth André
Leo Wanner

May 27–30, 2024

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Intelligent Interaction with Autonomous Assistants in the Wild

Organizers:
Yutaka Arakawa (Kyushu University, Japan)
Wolfgang Minker (Ulm University, Germany)
Elisabeth André (Augsburg University, Germany)
Leo Wanner (Pompeu Fabra University, Spain)

May 27–30, 2024

Abstract

This meeting gives an opportunity for experts from computer science, psychology, and ethics to collaborate and discuss the notion of "intelligent interaction" with autonomous assistants and the skills required for it. The main goals of the meeting are: To examine what "intelligent interaction" with autonomous assistants means and what capabilities are needed to achieve it. To discuss the requirements for intelligent interaction from multiple perspectives - computer science (natural language processing, multi-modal dialogue models, etc.), psychology (human factors, evaluating acceptance and trust), and ethics (privacy, social impact). To build interdisciplinary collaborations and jointly develop a research agenda for the field. To publish the meeting outcomes as open-access proceedings and use it as a steppingstone for joint publications, research collaborations, and funding applications across disciplines in the future. In essence, the meeting aims to bring together experts across domains to comprehensively explore the challenges and directions for realizing intelligent interaction with autonomous AI assistants through interdisciplinary discourse and cooperation.

1 Description of the Meeting

The early 21st century has witnessed a variety of software assistants integrating into our daily lives. Most prominently, there has been the emergence of speech assistants, such as Amazon Alexa and Google Assistant. These applications are frequently labeled as "intelligent" assistants. However, by human standards, they are mainly used for simple tasks, like playing music, providing weather updates, or switching lights on and off, while the interaction remains largely unsophisticated and command-based.

Recent advancements in autonomous agents capable of cooperating with users on complex tasks, such as problem-solving or decision-making, underscore the need to equip these systems with more advanced interaction skills to ensure they are accepted and trusted. Nevertheless, it remains unclear what constitutes intelligence in the context of interaction and what skills are required. These are the primary considerations of the proposed meeting.

In computer science, intelligence is commonly associated with artificial intelligence (AI), which is defined as the ability of machines to perceive their environment through sensory inputs and respond appropriately. Thus, machines attempt to emulate human cognitive functions, encompassing skills such as learning, planning, reasoning, adaptation, and natural language processing. Additionally, intelligent systems should leverage affective computing to interpret human emotions and act autonomously. Progress in computational processing power and machine learning has propelled research on autonomous assistants and robots, showing promising results in various application domains.

A major challenge in deploying this technology beyond laboratories is to create intelligent interactions. This task requires interdisciplinary collaboration across computer science, psychology, and ethics. Computer scientists face key challenges in developing technological models, methods, and strategies for communicating complex assistant functionalities. To achieve this, interactions must evolve from command-based exchanges to dynamic, reliable human-computer dialogues, initiated by either party and involving multiple participants when necessary. Utilizing multimodal sensory information, such as visual and physiological data, is crucial for intelligent interactions. Human factors and psychological models must also be integrated to enhance acceptance, trust, and usability. Ethical expertise is essential to address the privacy concerns and social impact of autonomous assistants.

An active collaboration among computer science, psychology, and ethics will provide a platform for exchanging ideas and benefiting from complementary expertise. We aim to foster an interdisciplinary dialogue and collaboration between experts from these fields.

This Shonan Meeting was designed to shorten the time for presentations and to allow more time for discussion and paper writing. Presentations were given in the PechaKucha style, in which each person had 200 seconds to introduce his or her research theme and what he or she wanted to discuss at this Shonan Meeting. All of them wormed up the audience by having them take a quiz that included two truths and one lie on the last page.

Discussions were held on the four topics of Theoretical Models, Technical Implementation, Evaluation and Ethics as perspectives to be considered in Autonomous Assistants in the Wild, and two groups were formed for each topic. Eight groups were formed, two for each topic, and discussions were held. Fi-

nally, the two groups were combined to write reports on the four perspectives. Since each report is several pages long, only a summary is given below. The full papers are attached at the end of this document.

Theoretical Models Group: This research group presents a theoretical model of an autonomous assistant capable of intelligent interaction with humans. The interaction cycles of perception, evaluation, goal, intention, and execution are discussed in combination to converge on actions and understandings shared by humans and autonomous agents. We will explore a framework for assistants that allows them to recognize multimodal inputs such as speech and gestures, represent knowledge, infer context, predict human biases, and manage common ground and shared goals with humans. We also seek to develop assistants that are comparable to humans in their ability to remember, manage cognitive load, make predictions, and model human behavior. Concepts such as affective models to simulate emotional states, personal and cultural models to understand a person's personality and cultural background, cognitive behavioral therapy for counseling, and the use of large-scale language models (LLMs) to simulate human-like interactions will be introduced. We also provide examples of using LLMs to simulate conversations that show emotions such as shame. In summary, we have presented a theoretical model and approach for developing autonomous assistants that can effectively interact with humans while understanding and adapting to their cognitive, emotional, and cultural characteristics.

Technical Implementation Group: Despite recent technological advancements in AI, many desired capabilities of intelligent assistant agents remain partially realized. Key intelligent interaction skills include situational and cultural awareness and adaptation, user needs identification and understanding, user collaboration to facilitate tasks, among others. In the wild, assistive agents should be able to navigate multi-party contexts, identify the relevant information for a certain context, and dynamically adjust behaviour depending on the current task, user emotions, role, initiative preferences, and so on. In this sense, we propose a co-creative approach, a combination of narrow AI solutions and virtual environment testing, to bridge the gap between current AI technology and human expectations.

Evaluation Group: The paper addresses the evaluation of intelligent interaction with autonomous agents, and more specifically what methodologies are suitable for evaluating their performance and impact on user experience in naturalistic settings. We argue for three key desiderata: 1) methodologies should infer changes in UX based on longitudinal data; 2) longitudinal data requires considering social and legal notions of acceptance; and 3) exploratory testing of IIAA behaviors should be limited by their specific deployment setting.

Ethics Group: This report is a discussion of **Ethical** aspects of Autonomous Assistants in the Wild, discussed at the Shonan Meeting No. 188, May 26-30, 2024 on Intelligent Interaction with Autonomous Assistants in the wild. Through two days of the meeting, the working group identified perspectives to consider in building and utilizing Autonomous Assistants. Paticularly, these regard 'dependencies on third parties', 'data privacy and protection', 'Biases and unintended consequences', 'cultural differences', 'legislation' as well as 'social impacts'.

2 Meeting Schedule

Time Table	Arrival Day	1st Day	2nd Day	3rd Day	Final Day
	May 26th	May 27th	May 28th	May 29th	May 30th
7:00 - 7:30					
7:30 - 8:00		Breakfast	Breakfast	Breakfast	Breakfast
8:00 - 8:30					
8:30 - 9:00					
9:00 - 9:30		Introduction	Plenary Working Session 2 (ChatGPT Design sprint	Plenary Working Session 4 (Debates)	Paper Writing Sessions
9:30 - 10:00		Pecha Kucha 1	Plenary Working Session 2 (ChatGPT Design Sprint	Plenary Working Session 4 (Debates)	Paper Writing Sessions
10:00 - 10:30		Break	Plenary Working Session 2 (ChatGPT Design Sprint	Break	Break
10:30 - 11:00		Pecha Kucha 2	Break	Paper Writing Sessions	Presentation of final results
11:00 - 11:30		Pecha Kucha 3	Working Groups	Paper Writing Sessions	Wrap up and Farewell
11:30 - 12:00		Working Group Allocation	Working Groups	Paper Writing Sessions	
12:00 - 12:30	Early check-in (negotiable)	Lunch	Lunch	Lunch	Lunch
12:30 - 13:00					
13:00 - 13:30		Group Photo Shooting		Excursion to Kamakura:	
13:30 - 14:00		Working Group	Plenary Working Session 3 (Ethical Dilemma Role-Play)		
14:00 - 14:30		Working Group	Plenary Working Session 3 (Ethical Dilemma Role-Play)		
14:30 - 15:00		Working Group	Break		
15:00 - 15:30	Regular check-in	Break	Working Groups		
15:30 - 16:00		Plenary Working Session 1 (Simulated Focus Groups)	Working Groups		
16:00 - 16:30		Plenary Working Session 1 (Simulated Focus Groups)	Group Keynotes		
16:30 - 17:00		Group Presentation of			
17:00 - 17:30		Preliminary Results			
17:30 - 18:00					
18:00 -18:30	Pre-meeting with Shonan staff	Dinner	Dinner		
18:30 - 19:00					
19:00 - 19:30	Welcome Banquet			Banquett	
19:30 - 20:00	Allocation of Chiefs				
20:00 - 20:30					
20:30 - 21:00					
21:00 - 21:30					
21:00 - 22:00					

Overview of Pechakucha Talks

In the first session on the first day, each participant described his/her research and expectations for this Shonan Meeting in Pechakucha style, in which 10 slides are explained in 20 seconds per page, for a total of 200 seconds.

Alaeddin Nassani

University of Aizu

Alaeddin Nassani currently serves as an Associate Professor at the University of Aizu in Japan, where he teaches virtual reality. His research interests span augmented reality (AR), virtual reality (VR), wearable computing, and human-computer interaction (HCI). Previously, he was a Research Fellow at the Auckland Bioengineering Institute, University of Auckland, focusing on research using digital humans for personal health management related to cardiovascular disease, diabetes, and mental health. At the Augmented Human Lab, he contributed to a national research project using electronic sensors to engage school kids in scientific inquiry. His work at the Empathic Computing Lab involved developing live 360 video streaming solutions for remote collaboration and tele-conferencing research.

Björn Schuller

TUM / Imperial College London

He received his diploma, doctoral degree, habilitation, and Adjunct Teaching Professor in Machine Intelligence and Signal Processing all in EE/IT from TUM in Munich/Germany where he is Full Professor and Chair of Health Informatics. He is also Full Professor of Artificial Intelligence and the Head of GLAM at Imperial College London/UK, co-founding CEO and current CSO of audEERING – an Audio Intelligence company based near Munich and in Berlin/Germany, Core Member in the Munich Data Science Institute (MDSI), Principal Investigator in the Munich Center for Machine Learning (MCML), Fellow of the Imperial Data Science Institute, and permanent Honorable Dean at TJNU/China and Visiting Professor at HIT/China amongst other Professorships and AUiliations. Previous stays include Full Professor and Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg/Germany (currently as Guest Professor), independent research leader within the Alan Turing Institute as part of the UK Health Security Agency, Guest Professor at Southeast University in Nanjing/China, Full Professor at the University of Passau/Germany, Key Researcher at Joanneum Research in Graz/Austria, and the CNRS-LIMSI in Orsay/France. He is a Fellow of the ACM, Fellow of the IEEE and Golden Core Awardee of the IEEE Computer Society, Fellow of the BCS, Fellow of the ELLIS, Fellow of the ISCA, Fellow and President-Emeritus of the AAAC, and Elected Full Member Sigma Xi. He (co-)authored 1,400+ publications (60,000+ citations, h-index=111 ranking him number 8 in the UK for Computer Science), is Field Chief Editor of Frontiers in Digital Health, Editor in Chief of AI Open and was Editor in Chief of the IEEE Transactions on AUective Computing amongst manifold further commitments and service to

the community. His 50+ awards include having been honoured as one of 40 extraordinary scientists under the age of 40 by the WEF in 2015. Currently, he was awarded IEEE Signal Processing Society Distinguished Lecturer 2024.

Christian Becker-Asano

Stuttgart Media University

He received his doctor's degree (Dr. rer. nat.) in Computer Science from the University of Bielefeld in 2008, for his work on affect simulation for agents with believable interactivity (WASABI architecture). He was Japan Society for the Promotion of Science (JSPS) pre-doctoral fellow in 2005 at the National Institute of Informatics, Tokyo, Tokyo, and JSPS post- doctoral fellow from 2008 to 2010 at ATR in Kyoto, Japan. In 2010 he became Junior Fellow at FRIAS in Freiburg, before in 2011 he joined the Research Group on the Foundations of AI at Freiburg University. From 2015 to 2020 he worked as researcher at Bosch R&D in Renningen (Stuttgart) and as Product Owner (Software, Bosch startup GmbH) in Ludwigsburg. In 2020 he became full professor at Stuttgart Media University the director of its newly founded Humanoid Lab.

David Traum

USCICT

He is the Director for Natural Language Research at the Institute for Creative Technologies (ICT) and Research Professor in the Thomas Lord Department of Computer Science at the University of Southern California (USC). He leads the Natural Language Dialogue Group at ICT. More information about the group can be found here: http://nld.ict.usc.edu/group/ Traum's research focuses on Dialogue Communication between Human and Artificial Agents. He has engaged in theoretical, implementational and empirical approaches to the problem, studying human- human natural language and multi-modal dialogue, as well as building a number of dialogue systems to communicate with human users. Traum has authored over 300 refereed technical articles, is a founding editor of the Journal Dialogue and Discourse, has chaired and served on many conference program committees, and is a past President of SIGDIAL, the international special interest group in discourse and dialogue. Traum earned his Ph.D. in Computer Science at the University of Rochester in 1994.

Elisabeth André

University of Augsburg

She received the degrees in computer science from Saarland University, including a doctorate. She is a full professor of computer science and founding chair of Human-Centered Multimedia with Augsburg University in Germany where she has been since 2001. Previously, she was a principal researcher with the German Research Center for Artificial Intelligence (DFKI GmbH) in Saarbrücken. She has a long track record in multimodal human-machine interaction, embodied conversational agents, social robotics, affective computing

and social signal processing. In 2010, she was elected a member of the prestigious Academy of Europe, the German Academy of Sciences Leopoldina, and AcademiaNet. To honor her achievements in bringing Artificial Intelligence techniques to HCI, she was awarded a EurAI fellowship (European Coordinating Committee for Artificial Intelligence) in 2013. In 2017, she was elected to the CHI Academy, an honorary group of leaders in the field of human-computer interaction.

Graham Wilcock

University of Helsinki

He is Adjunct Professor of Language Technology at University of Helsinki. He did his PhD at University of Manchester Institute of Science and Technology in 1999, after working in industry with ICL (International Computers Limited) at EU HQ in Luxembourg and with Sharp Corporation in Japan. He was co-organizer of several workshops on NLP and XML including the first Linguistic Annotation Workshop (LAW 2007). He received an IBM Innovation Award in 2008 for work on UIMA (Unstructured Information Management Architecture) and published a book Introduction to Linguistic Annotation and Text Analytics in 2009. Since 2015 he has worked on talking robots and conversational AI. He developed WikiTalk, a Wikipedia-based open-domain dialogue system for Nao robots, jointly with Kristiina Jokinen and they edited a Springer book Dialogues with Social Robots in 2017. He also developed CityTalk, a robot dialogue system using RASA conversational AI and knowledge graphs in Neo4j databases. In 2018-19 he was Visiting Professor at Kyoto University, where he worked in ERATO Ishiguro Symbiotic HRI project with the ERICA robot. He has presented robot demos at SIGDIAL 2015, COLING 2016, IJCAI 2018, IJCAI 2019 and research papers at RO-MAN 2021, HRI 2022, RO-MAN 2022, RO-MAN 2023, HRI 2024, IWSDS 2024.

Ines Lobo

GAIPS, INESC-ID, IST

She is a Ph.D. student in the AI for People and Society Research Group at INESC-ID, Instituto Superior Técnico, Lisbon, Portugal. She has a master's degree in Information Systems and Computer Engineering, specializing in Games and Intelligent Systems.

Inés Torres

University of the Basque Country

She received her PhD in Physics from the UPV/EHU in 1990, including an internship at the CNET- Lanion (France). She was a visiting researcher at the Polytechnic University of Valencia (Spain) (1991-1992), visiting Faculty in Carnegie Mellon University (USA) (20212) and visiting Professor at the University of California (UCSC) under the Fulbright program (2018). She is currently a Full Professor of Computer Science at the UPV/EHU. Prof. Torres has a

multi-disciplinary academic and industrial background in the fields of Speech and Language Technologies focusing on data-driven approaches. Her current research interests involve Human-Machine Interaction, Emotional Speech Processing, Spoken Dialogue Systems and their applications. She has supervised 11 PhD students, and she currently supervises two more. Prof. Torres has recently coordinated the H2020 EMPATHIC project, led UPV/EHU's participation in the H2020- MSCA-RISE MENHIR, was a member of the Scientific Advisory Board for the e-VITA EU-Japan Virtual coach for smart ageing project and published in outstanding scientific journals and conferences, among other activities such as national projects, industrial research or research contracts with companies.

Jesse Thomason

USC Viterbi School of Engineering

He is an Assistant Professor at USC leading the Grounding Language in Multimodal Observations, Actions, and Robots (GLAMOR) lab. Language is not text data, it is a human medium for communication. The natural language processing (NLP) community at large has doubled down on treating digital text as a sufficient approximation of language, scaling larger datasets and corresponding models to fit that text. He has argued that experience in the world grounds language, tying it to objects, actions, and concepts. In fact, he believes that language carries meaning only when considered alongside that world, and that the zeitgeist in NLP research currently misses the mark on truly interesting questions at the intersection of human language and machine computation. His research enables agents and robots to better understand and respond to human language by considering the grounded context in which that language occurs. His research weaves together three core threads: 1) learning with language, perception, and action; 2) neurosymbolic reasoning for language, vision, and robotics; and 3) language processing for accessibility and health. His lab has received funding from the Defense Advanced Projects Research Agency (DARPA), the National Science Foundation (NSF), the National Institute of Health (NIH), the Army Research Laboratory (ARL), and the Laboratory for Analytical Sciences (LAS).

Kaoru Sumi

Future University Hakodate

She is a professor in Future University Hakodate, Japan. Prof. Sumi received her Ph.D. in engineering from the University of Tokyo. Prior to joining academia, he had eight years of industry experience at KDDI and Mitsubishi Corporation. She previously worked at ATR MI&C Research Laboratories, Communications Research Laboratory, and Osaka University, where she researched human-computer interaction, knowledge engineering, and the application of artificial intelligence. After Prof. Sumi worked on media informatics and human-agent interaction at the National Institute of Information and Communications Technology (NICT), and Hitotsubashi University. She was a visiting professor in British Columbia, Canada.

Kristiina Jokinen

AI Research Center, National Institute of Advanced Industrial Science and Technology (AIST)

She is a Senior Researcher at AI Research Center (AIRC) at National Institute of Advanced Industrial Science and Technology (AIST) in Tokyo Waterfront, and Adjunct Professor at University of Helsinki. She is also an Advisory Board Member for the AI in Engineering Programme in Japan, and for the IWSDS series of Dialogue Workshops. She is Life Member of Clare Hall at University of Cambridge, and Member of the European ELLIS network. Her first degree in physics was at the University of Helsinki, and she received her PhD from UMIST, University of Manchester. She was awarded a JSPS Fellowship for postdoc research at NAIST (Nara Institute of Science and Technology), after which she was Invited Researcher at ATR Research Labs in Kyoto, and Visiting Professor at Doshisha University. Her research concerns cooperative human-robot interaction, AI-based dialogue modelling and multimodal communication. She has published widely on these topics, including three books. She developed Constructive Dialogue Model as a general framework for interaction, and together with Graham Wilcock she developed the Wikipedia-based robot dialogue system WikiTalk, which won the Special Recognition for Best Robot Design (Software Category) at the International Conference of Social Robotics in 2017. She has led numerous national and international research projects and is currently leading dialogue research for a trustworthy virtual coaching application in the large EU-Japan collaboration project e-VITA.

Leo Wanner

Pompeu Fabra University

He holds a Master (Diploma) degree in Computer Science from the University of Karlsruhe and a PhD in Computational Linguistics from the University of The Saarland, Germany. Since 2005, he has been ICREA Research Professor at the Pompeu Fabra University in Barcelona. Before joining ICREA, Leo held research positions at the German National Center for Computer Science, University of Waterloo (Canada), University of Stuttgart, and University Pompeu Fabra. As visiting researcher, he was also affiliated with, among others, the Information Sciences Institute of the University of Southern California, Columbia University, and University of Augsburg. Leo published 240+ peer reviewed papers and edited 10 volumes in different areas of Computational Linguistics. He is member of the International Committee for Computational Linguistics (ICCL), Associate Editor of the Computational Intelligence and Frontiers in AI, Language and Computation journals, and serves as regular reviewer for a number of high-profile conferences and journals on Computational Linguistics. Throughout his career, Leo worked on a number of topics in the field, including multilingual conversational agents, natural language generation and summarization, concept extraction, and hate speech recognition.

Matthias Kraus

University of Augsburg/LMU

He is currently a Post-Doc at the Chair of Human-Centered Artificial Intelligence at Augsburg University and has been an interim professor at LMU Munich. He has extensive expertise on various fields within HCI, HRI as well as AI, such as social robotics, multi-modal interaction, user modeling, situation- and user-adaptive dialogue management, and natural language processing. He has published +35 papers at top-tier venues in AI, HMI, and HRI as well as contributed to several book chapters. Furthermore,his work has been carried out within large-scale national and international projects considering the integration of proactive behavior in cognitive assistants and social robots in work-related and private contexts. He has published and collaborated with several groups at other national and international universities and industrial research companies, including the University of Granada, TU Eindhoven, Robert Bosch GmbH, and KUKA GmbH.

Matthias Rehm

Aalborg University

He is head of the Human Machine Interaction group at the Technical Faculty of IT and Design at Aalborg University in Denmark. He is also the director of the interdisciplinary Laboratory for Human Robot Interaction at Aalborg University (https://hri.tech.aau.dk). He received his Diploma and Doctoral degrees (with honors) in 1998 and 2001 respectively from Bielefeld University in Germany. In 2008, he successfully completed his habilitation process in Informatics at the University of Augsburg in Germany. His research is focused on modeling social, affective and cultural aspects of everyday behavior for intuitive human machine interactions. He has over 150 peer reviewed publications in the area of Robotics, HCI, Technology Enhanced Learning, Multimodal Interaction, and Culture Aware Technology. In 2010, he became founding and steering group member of Aalborg University's cross- departmental robotics program Aalborg U Robotics (http://robotics.aau.dk). From 2015 to 2019 he was the elected vice president for the International Association for Smart Learning Ecosystems and Regional Development (http://aslerd.org).

Michael Cohen

Aizu Universitu

He is Prof. Emeritus at the U. of Aizu in Aizu-Wakamatsu, Fukushima. He received an Sc.B. in EE from Brown University (Providence, Rhode Island) in 1980, M.S. in CS from the University of Washington (Seattle) in 1988, and Ph.D. in EECS from Northwestern University (Evanston, Illinois) in 1991. He has worked at the Air Force Geophysics Lab (Hanscom Field, Massachusetts), Weizmann Institute (Rehovot; Israel), Teradyne (Boston, Massachusetts), BBN (Cambridge, Massachusetts and Stuttgart; Germany), Bellcore (Morristown and Red Bank, New Jersey), the Human Interface Technology Lab (Seattle, Washington), and the Audio Media Research Group at the NTT Human Interface

Lab (Musashino and Yokosuka; Japan). He has research interests in interactive multimedia, including extended reality (XR), computer music, spatial audio & stereotelephony, stereography, ubicomp, and mobile computing. He is on the Scientific C o mmittee o f t h e J o urnal o f V i rtual R e ality a n d B r oadcasting, on the Advisory Board of the Int. J. of Applied and Creative Arts and an Assoc. Editor of Presence: Virtual and Augmented Reality. He is a member of the ACM, IEEE Computer Society, 3D-Forum, TUG (TeX Users Group), and VRSJ (Virtual Reality Society of Japan).

Nils Mandischer

University of Augsburg

Nils Mandischer is a post-doc at the University of Augsburg in the Chair of Mechatronics. His work is in the field o f h uman-robot t e aming. H i s particular research interests are the assessment of human capabilities towards increasing the level of autonomy of robotic assistants. His works covers perception and autonomy methods in the domains of robotic assistants for people with disabilities and collaborative rescue robotics.

Sebastian Zepf

Mercedes-Benz AG

He is an HCI Researcher at Mercedes-Benz AG, working on designing and developing AI assistants that anticipate user needs and proactively initiate actions for and interactions with users whenever suitable. Before that, he led the User Experience and Usability Engineering team at B. Braun New Ventures GmbH, a corporate start-up that aims to digitize and automate neurosurgical procedures. Sebastian obtained his PhD in Computer Science in 2021 from Ulm University, in collaboration with the Mercedes-Benz AG and the MIT Media Lab. His research interests includes proactive and multimodal interaction, personalized user interfaces, affective c omputing, a nd u ser s ensing a nd modeling.

Seitarou Shinagawa

SB Intutions

He graduated from the Faculty of Engineering at Tohoku University in 2013. He completed his Master's degree in Information Science at the same university in 2015, and his doctoral studies at the Nara Institute of Science and Technology in 2020, earning a Doctor of Engineering. After serving as an Assistant Professor at the same university, he joined SB Intuitions Corp. in 2024, where he is engaged in the research and development of multimodal foundation models.

Serge Thill

Radboud University

He is an associate professor in artificial intelligence and principal investigator at the Donders Centre for Brain, Cognition, and Behaviour at Radboud University Nijmegen (Netherlands). I am the chair of the newly established department

of Human Centred Intelligent Systems (HuCIS) and I lead the Foundations of Intelligent Technology (FoundIT) research group. I am also part of the National Education Lab AI (NOLAI, www.nolai.nl), a National Growth Funds- funded initiative on the use of technology an AI in Dutch primary and secondary education, where I lead the scientific team on technological aspects of AI. I am also co-editor in chief for the journal Cognitive Systems Research and associate editor at Adaptive Behaviour and the International Journal of Social Robotics.

Shin Katayama

Nagoya University

He is a Project Assistant Professor at the Graduate School of Engineering, Nagoya University, Japan. He earned his Doctor of Engineering degree from Nagoya University in 2023. Prior to this, he received a Master of Media and Governance degree and a Bachelor of Arts in Environmental Information degree from Keio University. His research interests focus on Human-Computer Interaction, Dialogue Systems, and Affective Computing.

Stephan Sigg

Aalto University

Stephan Sigg is an Associate Professor at Aalto University in the Department of Information and Communications Engineering. His research interests include the design, analysis and optimisation of algorithms for distributed and ubiquitous systems. Especially, his work covers proactive computing, distributed adaptive beamforming, context-based secure key generation and device-free passive activity recognition. Stephan is an editorial board member of the Elsevier Journal on Computer Communications and has been a guest editor for the Springer Personal and Ubiquitous Computing Systems Journal. He has served on the organizing and technical committees numerous prestigious conferences including IEEE PerCom, ACM Ubicomp.

Susanna Pirttikangas

University of Oulu

Research director Pirttikangas has extensive background in artificial intelligence related research and business. She has served as a member in the Ministry of Economic Affairs and Employment of Finland AI4.0 technology leadership group accelerating the development and introduction of artificial intelligence (AI) and other digital technologies in com- panies. This work is continuation of the work under Finland's AI program (AI era) Data and Platform Economy working group chaired by Kimmo Alkio, Tieto and AI accelerator Task Force chaired by Pekka Ala-Pietilä. She also works as a freelancer lead AI scientist in a Finnish SME Silo.AI and runs her own company Tausta Oy, providing artificial intelligence services. She frequently educates companies and research organizations on AI related state-of-the-art technologies.

Tadashi Okoshi

Keio University

He is an associate professor in Faculty of Environment and Information Studies, Keio University. He is a computer scientist especially focusing on information and computing systems for supporting our life-long wellbeing. His major is mobile and ubiquitous computing, context-aware computing etc. His recent research works are on human attention management, mobile affective computing, and computing for well-being (WellComp). He has served as organizing and program committee member of mobile and ubiquitous systems, and networking conferences and workshops. He sits on the editorial boards of ACM Proceedings on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT). He has been serving as social media director of ACM SIGMOBILE since 2016. In 2019, he was awarded IPSJ Microsoft Faculty Award, an annual award for young researchers who have made outstanding international contributions to research and development in major areas of informatics. He holds B.A. in Environmental Information (1998), Master of Media and Governance (2000) from Keio University, M.S. in Computer Science (2006) from Carnegie Mellon University, and Ph.D. in Media and Governance (2015) from Keio University, respectively. He also has over 7-year experiences of entrepreneurship, software architecting, product management, and project management in IT industries (Web2.0, blogging, social networking and social media).

Tsunenori Mine

Kyushu University

He is an Associate Professor at Department of Advanced Information Technology, Faculty of Information Science and Electrical Engineering, Kyushu University. His research interests include developing real services using artificial intelligence techniques, in particular, natural language processing, text mining, data mining, recommendation, and multi-agent systems. He pays particular attention to representation learning when developing certain methods in his research. He is currently leading several joint research projects with several companies and academic institutions to develop technologies and theories that are both practical and academically novel. He serves as a reviewer and PC member for top journals and conferences in his field such as IEEE transaction on Intelligent Transportation Systems, AAAI, AAMAS, AIED, ECML-PKDD, PAKDD.

Wolfgang Minker

Ulm University

He eceived his Diploma (M.Sc.) in Engineering Science from University of Karlsruhe, Germany. He finished his Ph.D. in Engineering Science at the University of Karlsruhe, Germany in 1997 and received a Ph.D. in Computer Science from Université Paris-Sud, France in 1998. From 1993 until 2000 he was a teaching assistant at LIMSI-CNRS, Université Paris-Sud, France and subsequently worked as Senior Researcher at the Dialogue Systems Group of

DaimlerChrysler Research and Technology, Ulm, Germany from 2000 to 2003. Since 2003 he is a full professor at Ulm University, Germany.

Yasuyuki Sumi

Future University Hakodate

He has been working on enabling knowledge creation between people and between people and agents based on conversations. In this short talk, he introduced his related projects to date, such as, multi-modal measurement and understanding multiparty conversations, context-aware digital assistant, nonverbal analysis of tutoring dialogues, facilitating knowledge circulation by embedding conversations in real-world situations, and measuring social activities by lifelog.

Yuki Matsuda

Okayama University

Yuki Matsuda is a Lecturer at the Faculty of Environmental, Life, Natural Science and Technology, Okayama University. He has been working on intelligent navigation and coaching systems in the wild by combining sensing technology that includes the emotions and behaviors of users, and dialogue technology to generate natural conversation between AIoT (artificial intelligence of things) and people. He has introduced ongoing projects for tourism navigation, museum guidance, and Japanese abacus coaching.

Yutaka Arakawa

Kyushu University

I introduced my research topic, human activity recognition. With the evolution of smartphones and wearables, mobile devices are now able to sense human movements anytime and anywhere. In addition, push notifications allow us to intervene at any time, and in recent years, applications such as SaMD (Software as a Medical Device) and DTx (Digital Theraputics) have been put to practical use. Technologies to change human behavior will continue to play an important role in human society, and the intelligent interaction technologies used in such technologies will continue to be an important field of research. Currently, there is a wide range of collaborative research on behavior change support, the most significant of which is the joint project with Moomin Move. At this Shonan Meeting, we will discuss "autonomous assistants" that contribute to behavior change, How to integrate the results of "behavior recognition' into dialogue generation and adaptive and proactive intervention? Also, how to adapt the contents of dialogue for each user's individual? How to take "personality" of the user and "environmental context" into consideration?

List of Participants

- Alaeddin Nassani, Aizu University
- Bjoern Schuller, TUM / Imperial College London
- Christian Becker Asano, Stuttgart Media University
- David Traum, USC ICT
- Elisabeth André, University of Augsburg
- Graham Wilcock, University of Helsinki
- Ines Lobo, GAIPS, INESC-ID, IST
- Inés Torres, University of the Basque Country
- Jesse Thomason, USC Viterbi School of Engineering
- Kaoru Sumi, Future University Hakodate
- Kristiina Jokinen, AIST
- Leo Wanner, Pompeu Fabra University
- Matthias Kraus, University of Augsburg
- Matthias Rehm, Aalborg University
- Michael Cohen, Aizu Universitu
- Nils Mandischer, University of Augsburg
- Sebastian Zepf, Mercedes-Benz AG
- Seitarou Shinagawa, SB Intutions
- Serge Thill, Radboud University
- Shin Katayama, Nagoya University
- Stephan Sigg, AaltoUniversity
- Susanna Pirttikangas, University of Oulu
- Tadashi Okoshi, Keio University
- Tsunenori Mine, Kyushu University
- Wolfgang Minker, Ulm University
- Yasuyuki Sumi, Future University Hakodate
- Yuki Matsuda, Okayama University
- Yutaka Arakawa, Kyushu University

3 Summary of Discussions

In the following the results of the workshop's discussion are summarized in topic-related sections.

3.1 Theoretical Models for Autonomous Assistants

3.1.1 Introduction

The following paper is structured hierarchically into meta-models (Section 3.1.2), components of enabling autonomous assistance (Section 3.1.3), cognitive and social models for LLMs (Section 3.1.4), and multi-agent systems (Section 3.1.5).

3.1.2 Meta-Models of Human-Autonomy Interaction

Meta-models describe an abstracted sequence of actions which need to be fulfilled to foster a good interaction between an autonomous agent and the human. There exist many models from diverse domains which lay different focus on the type of interaction and the interacting agents. In this section, we bring together some of those models to foster a common understanding of autonomous and intelligent interaction. Therefore, we first introduce the interaction cycle and arbitration (Secion 3.1.2), before combining both into a novel meta-model (Section 3.1.2). Even though all presented meta-models are centered about dyadic interaction, all may be extended towards multi-agent interaction.

To generate personalized motivational messages, self-determination theory [15] has been widely used. It mentions that human motivation is driven by three psychological needs; autonomy, competence and relatedness. Autonomy is the desire to feel in control of actions. Competence involves mastering tasks and learning new skills. Relatedness is the need for connection and belonging. These needs can be intrinsic or extrinsic motivations.

The design of AI system can benefit from this theory to enhance motivation and engagement. Agents can support autonomy of users to personalise their actions interactions and choices. For competence, agents can provide guidance to users to achieve their goals. For relatedness, agents can be empathetic and socially supportive. Related Meta-Models This section discusses a few typical meta-models for autonomous interaction. Section 3.1.2 introduces Norman's interaction cycle and its main ideas. Section 3.1.2 discusses arbitration and its separation of the autonomous assistant into the technical system and autonomous agent.

Interaction Cycle One of the most fundamental models outlining the process the user goes through when interacting with a system is the "interaction cycle" described by Norman [45]. The cycle highlights that it is crucial to understand user needs and focus on usability in system design dividing the interaction process into two central phases as depicted in figure 1: The execution and the evaluation. The execution phase contains the user's internal process of forming his/her goal and intended action as well as the concretization on how to perform the action, ending with the actual execution of this action. The evaluation phase on the other hand contains the users processing of the system response, covering the perception and interpretation of the system response as well as

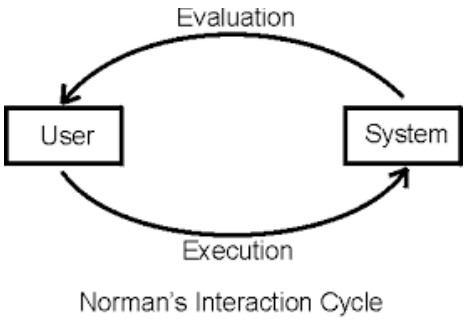

Figure 1: The basic interaction cycle by Norman.

the evaluation of the outcome regarding interaction success with respect to the given goal.

Arbitration Cycle In shared autonomy, arbitration is the interaction paradigm, which allows two agents to converge on a consensual control decision. Arbitration is linked with the H-metaphor, that symbolizes the autonomy as a horse and the human as a rider. In the model, there are three entities:

- Human Agent
- Autonomous Agent
- Technical System

In contrast to Section 3.1.2, within this model the autonomous agent and the technical system are separated, even though many technical systems directly embed the autonomous agent, e.g., in an autonomous car or in a smart home assistant. Figure 2 depicts the arbitration cycle. Within, the human and autonomous agent interact in order to solve a dissent (or dissonance) into a consensual decision (consonance). Meanwhile both perceive the states of the technical system and the other agent(s) and control the technical system (here: a car) according to their individual intent. This interaction system may feature dyadic or multi-agent interactions.

On a side note, the human may also be considered an autonomous agent, even though their nature is not artificial but anthropomorphic.

System, Interaction System, and Technical System To clarify on the following thoughts, it is important to note which system is addressed. The *technical system* is the technical entity that is to be influenced, controlled or acted on by all involved agents – according to Section 3.1.2. The *system* contains everything within system boundaries including agents, technical system, and potentially the context and environment. The *interaction system* involves only the part of the *system* that is relevant for the interaction between the agents.

Combining the Cycles We can consider combining the interaction and arbitration cycle as shown in Figure 3. The left indicates the arbitration cycle and the right represents the interaction cycle. Here, we consider only two agents:

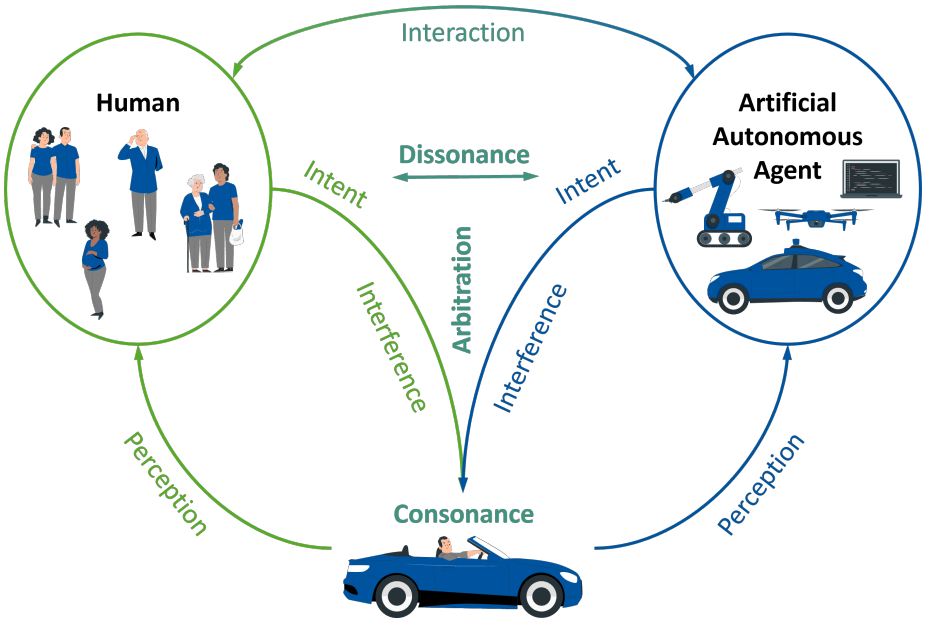

Figure 2: Arbitration model, based on Flemisch et al. [20].

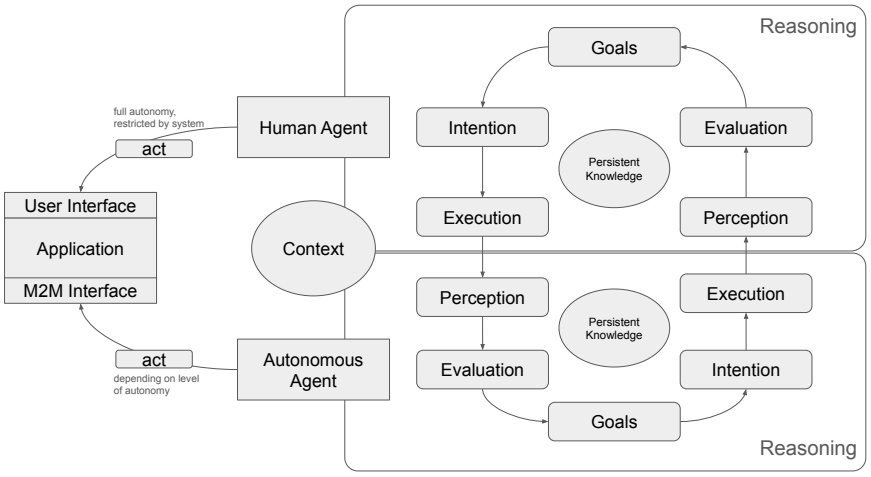

Figure 3: The combined cycle of interaction and arbitration.

Human and Autonomous Agents in the environment for simplification, but of course, we can extend the cycles to multi-human and multi-agent communication.

The Human and Autonomous Agents share the application and they can affect it through the interfaces for each. For example, in autonomous driving, the human driver can handle the car and the autonomous driving system can also control it to a certain extent, to achieve the shared goal, i.e., arriving at the

destination. The privilege level of control depends on the level of autonomy. In a fully autonomous environment, human actions may be restricted depending on the application.

The Human and Autonomous Agents have the same reasoning process indicated by the mirrored cycles: the Perception module receives the multimodal inputs and the contexts, then the Evaluation module judges the current states. The Goals module handles the planning and optimizes the current plan to achieve the goal. According to the planning, the Intention module assesses an Agent intention to act. Finally, the Execution module demonstrates multimodal output and performs an agent action. The interaction between the Human and Autonomous Agent consists of a closed loop of their reasoning process. In the reasoning process, each brings their persistent knowledge, and the knowledge is reinforced through persistent interaction.

3.1.3 Components of Human-AI Companionship

In this section, we introduce the components required for a successfull autonomous interaction. Therefore, we structure the components into the human (Section 3.1.3) and autonomous agent (Section 3.1.3).

Human Agent In this section, we list and summarize theories and models around aspects of human cognition that should be considered for deriving a meta-model of an intelligent interactive autonomous assistant that is capable of giving support to a user in given applications and tasks that equals or exceeds human performance.

The first relevant aspect is human information processing and memory. Atkinson and Shiffin [4] propose a set of three different components to model human memory:

1. Sensory Register: This component is the initial step of information processing that buffers sensory information from the human senses for a very short duration (milliseconds to seconds) before initiating further processing.

2. Short-Term Memory: The short-term memory holds information from the sensory register that is given attention to and maintains this information for 15 to 30 seconds. Rehearsel processes can help to keep information in the short-term memory for a longer duration.

3. Long-Term Memory: This component can store information that was encoded form the short-term memory for a much longer time and provides much higher capacity. To retrieve information from this knowledge base back to the short-term memory, retrieval processes are necessary.

Overall, this model highlights the importance of control processes and the importance of managing information flow between the memory components for enhancing the retention of memory and retrieval.

Another theory of the human mind, the so-called dual-system theory, considers two different modes of thinking. The first mode is responsible for instinctual judgements and operates quickly, requires very little effort, and does not require voluntary control. The second mode includes effortful mental activities,

is thus slower and requires conscious effort [27]. It is shown that individual differences in the interplay of the two modes influence cognitive abilities such as reasoning [58].

Focusing on the nature of working memory, Sweller [60] introduced the cognitive lead theory, which posits that the cognitive capacity of working memory is limited. Within the theory, it is differentiated between problem-solving processes and learning processes, with problem-solving inducing a very high cognitive load and thus being detrimental to learning.

The Adaptive Control of Thought-Rational (ACT-R) is a theory presenting a cognitive architecture aiming to model the human mind as a system including modules such as memory, perception, and action [2]. The theory details how the different modules interact with each other and can be used to simulate various cognitive tasks such as problem-solving and learning. In his book "How can the human mind occur in the physical universe?", Mellon [40] also highlights how these cognitive tasks can be studied through computational modeling.

Finally, the aspect of embodied cognition argues that the traditional cognitive theories are incomplete, describing importance of the deep interconnection between cognitive processes and sensory, motor, and affective systems [6]. In addition, Wilson [69] examines six different perspectives within embodied cognition such as different situations and time-dependency, highlighting the complexity of embodied cognition and the importance of considering the human body in cognitive processes.

Autonomous Agent To fulfill the individual steps in interaction, the autonomous agent must implement certain components:

1. Perception of multimodal inputs (e.g., speech, gestures, body motion)
2. Context awareness
3. Input understanding (e.g., Natural Language Understanding, context changes)
4. Reasoning & problem-solving
5. Intrapersonal and interpersonal intelligence (analogous to Gardner)
6. Multimodal interaction management, including transparency of own behavior
7. Knowledge representation & retrieval
8. Reflection & learning
9. Common ground & shared goal model
10. Predictive coding & convergence on actions
11. Model of human biases (e.g., embodied cognition)

First, the agents needs to perceive their surroundings given by multimodel inputs (1), particularly natural inputs like language or body behavior. This is connected with the understanding of these inputs (3) to form the context (2), which again establishes one form of context awareness of the agent. To properly perceive these information, the agent needs to consider human biases (11) and

Human Equivalent	Autonomy Equivalent
theory of mind	common ground, shared goal model
sensory memory, human senses	input understanding, multimodality
short-term memory, long-term memory	knowledge representation
information processing, problem-solving, intelligence	context awareness, reasoning, problem-solving, intrapersonal intelligence, interpersonal intelligence
congitive load, memory retrieval	knowledge retrieval
embodied cognition	model of human biases
anticipation	predictive coding

Table 1: Human and autonomy equivalent of their individual components.

feature some kind of perception model of the human. From all actions, knowledge is generated and stored (7). The agent then uses knowledge and context for reasoning (4), i.a., to solve problems encountered during interaction (see Section 3.1.2). When the agent is confronted with "problems", it must reflect on the nature of these problems, which leads to learning (8). Learning is an aspect of intelligence. We project the autonomous agent to explicitly having a need for interpersonal and intrapersonal intelligence (5), given its role as an intelligent assistant. To finally come to a conclusion of self-action, the agent uses a shared goal model (9) to reason on the shared goal and potentially diverging own goals and converges on an action while predicting human behavior (10). To this end, a multimodal interaction manager (6) is used. To counteract misunderstanding in interaction, the autonomous agent needs to employ transparent actions, which are understandable for the human. It may also be considered that the human learns the robot behavior and adapts towards the autonomous agent, similar to how the autonomous agent adapts its behavior towards the human. Given such dyadic adaptability, transparent actions would support the learning process but would be less important than in an interaction system where the human does not at all adapt towards the autonomous agent.

What becomes obvious is that the optimal autonomous agent needs to become equivalent or better than the human in all abilities. This is indicated by many components having an equivalent in both agents. The only component that partially falls out is the dual-process theory. Even though in theory, the autonomous agent could also simulate both procedures of processing information, namely fast-automatic-intuitive[1] and slow-deliberate-analytical, to use it to full extent, the automation shall act fast-automatic-analytic. This is an ability that the human cannot have and where the autonomy exceeds human abilities.

3.1.4 Adapting Cognitive and Social Models to LLMs

In this section, we introduce the theory of the emotional model, the personal model and the cultural model and provide representative examples of theories from the cognitive and social sciences that have been widely used as a foundation to develop virtual coaches.

[1]To simulate intuitive actions, the autonomous agent would still base its decisions on analytical processes, but solved faster than the human is able to. Hence, the fast-automatic-intuitive (human) converges with fast-automatic-analytic (autonomy).

Theories of Affect and Personality In human-system interaction, it is useful to be able to recognize human emotions, situations and background in real time and react to them immediately. Here, we introduces the emotional model, the personal model and the cultural model.

Emotional Models Emotional modeling involves basic emotion models, dimensional models, emotional vector models. In the basic emotion models, there are Ekman's model [17] and Plutchik's Wheel of Emotions[49]. Dimentional models includes Russell's Circumplex Model of Affect and Pleasure-Arousal-Dominance Model [53] . The Ortony, Clore, and Collins (OCC) model [46] is a well-known framework in the field of artificial intelligence and psychology that describes the cognitive structure of emotions. The OCC model categorizes emotions based on the cognitive appraisal of events, agents, and objects. Some studies use Deep Learning (Convolutional Neural Network (CNN) [50] [72] , Recurrent Neural Network (RNN) [24] [57] , Tranformer [16] [59]) and Sentiment Analysis (NLP) [35] [61] .

Personal Models Systems that tailor interactions to the personality and cultural background of the system's users can be effective in their respective domains. The Big Five Theory is a hierarchical model of personality traits that categorizes them into five major dimensions: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. [39] The Myers-Briggs Type Indicator (MBTI) [43] is based on Carl Jung's theory of psychological types. It classifies individuals into 16 personality types based on four dichotomies: Extraversion/Introversion, Sensing/Intuition, Thinking/Feeling, and Judging/Perceiving. Eysenck's Three-Factor Model [18] , proposed by Hans Eysenck, includes three dimensions: Extraversion, Neuroticism, and Psychoticism. This model emphasizes the biological basis of personality traits. The HEXACO Model [3] is an extension of the Big Five, including a sixth dimension: Honesty-Humility. It addresses the role of honesty and humility in social behavior. Cattell's 16 Personality Factors [12] were identified through factor analysis, resulting in 16 primary personality factors. There are genetic and biological approaches, such as research into the heritability of personality traits and the influence of genetics, as well as the exploration of the neural and hormonal underpinnings of personality [9] [66] .

Cultural Models Several theories investigate how cultural backgrounds influence human behavior, communication, and organizational practices. These theories often explore how cultural norms and values shape individuals' actions and interactions. Culture Map [41] is a tool to understand and navigate cultural differences in a global business environment. The framework focuses on eight dimensions of cultural variability: Comminicating, Evaluating, Persuading, Leading, Deciding, Trusting, Disagreeing, and Scheduling. Hofstede's Cultural Dimensions Theory [25] was developed by Geert Hofstede, this theory identifies six dimensions that describe national cultures: Power Distance, Individualism vs. Collectivism, Masculinity vs. Femininity, Uncertainty Avoidance, Long-Term vs. Short-Term Orientation, and Indulgence vs. Restraint. These dimensions help explain how cultural values influence behavior in different societies. Hall's Cultural Context Theory [22] distinguishes between high-context

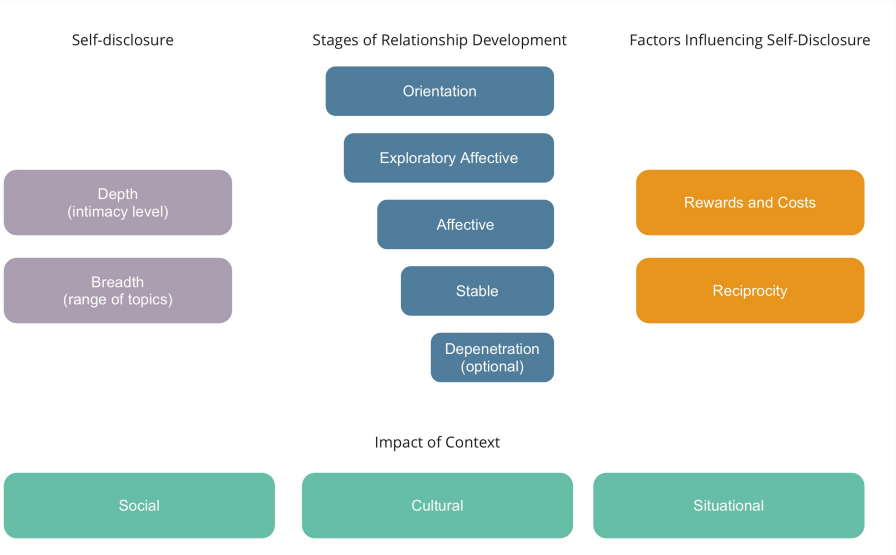

Figure 4: Social Penetration Theory

and low-context cultures. High-context cultures rely heavily on implicit communication and nonverbal cues, whereas low-context cultures depend on explicit verbal communication. Trompenaars' Seven Dimensions of Culture [23] includes seven dimensions: Universalism vs. Particularism, Individualism vs. Communitarianism, Specific vs. Diffuse, Neutral vs. Emotional, Achievement vs. Ascription, Sequential vs. Synchronic Time, and Internal vs. External Control. These dimensions help understand how cultural differences impact business and management practices. The Global Leadership and Organizational Behavior Effectiveness (GLOBE) study [26] examines cultural influences on leadership and organizational behavior. The study identifies nine cultural dimensions: Performance Orientation, Assertiveness, Future Orientation, Humane Orientation, Institutional Collectivism, In-Group Collectivism, Gender Egalitarianism, Power Distance, and Uncertainty Avoidance.

Cognitive Behavior Therapy Cognitive behavioral therapy (CBT) focuses on recognizing and changing negative patterns of thought and behavior. It combines principles of cognitive therapy (which deals with thoughts and beliefs) and behavioral therapy (which focuses on changing behaviors). CBT aims to help individuals understand the relationship between their thoughts, feelings and behaviors and to develop coping and change skills.

Virtual agents in the role of counselors frequently use cognitive behavior therapy a a guideline to structure the conversation with a human. Examples include virtual agents that interact with patients suffering from depression, mental disorders, or anxiety [29]. Usually, the dialogues are scripted based on data collections with clinical dialogues.

Social Penetration Theory The social penetration theory [1] (Figure 4) provides a framework for understanding the development of interpersonal re-

lationships through the process of self-disclosure. There are different levels of connections from basic to deeper levels. Self-disclosure can increase or decrease based on breadth (range of topics) and depth (intimacy of information shared) as the relationship progresses. Factors influencing the self-disclosure are the perceived rewards and costs, reciprocity and the context. The context can be sub-divided into social, cultural and situational context. The relationship development can be more dynamic and non-linear. The relationship can regress with reduced self-disclosure leading to less connection. Mutual self-disclosure increase connection and apply to human-agent relationship.

[13] reviewed the theories of emotional bonds formed between users and conversational agents. They explored the development and impact of social companionship with AI-enabled conversational agents, emphasizing their role in providing emotional support and building consumer relationships. It identifies key trends and constructs in the domain, offering a conceptual framework that includes antecedents, mediators, moderators. They also explored the ethical implications and future research directions, emphasizing the need for a macroscopic view to guide the design of efficient and ethical AI companions

Self-Determination Theory To generate personalized motivational messages, self-determination theory [15] has been widely used. It mentions that human motivation is driven by three psychological needs; autonomy, competence and relatedness. Autonomy is the desire to feel in control of actions. Competence involves mastering tasks and learning new skills. Relatedness is the need for connection and belonging. These needs can be intrinsic or extrinsic motivations.

The design of AI system can benefit from this theory to enhance motivation and engagement. Agents can support autonomy of users to personalise their actions interactions and choices. For competence, agents can provide guidance to users to achieve their goals. For relatedness, agents can be empathetic and socially supportive.

Combining Theories and Signal Processing MARSII is a model of appraisal, emotion regulation, and social signal interpretation [21]. By coming theories of affect with machine learning approaches, it enables us to infer affective states both from their causes and their expressions. To illustrate the idea, let's have a look at Figure 5 which has been designed using recordings of job interviews. Imagine a user is told by the job interviewer that he is not appropriately dressed. How would a user feel in such a situation? By running a simulation, we might come to the conclusion that the user feels shame. Typically, shame is reflected by blushing, touching the face, avoiding eye contact etc. However, in that particular case, the module for social signal interpretation detects instead a sequence of smiles which leads us to the conclusion that the interviewee is regulating his shame instead of showing it directly. According to Nathanson's Compass of Shame, there are various ways to regulate shame. By combining the findings from the affect simulation and the social signal analysis, a system might come to the conclusion that the interviewee most likely applies the regulation strategy "Avoidance" meaning that the interviewee applies strategies to distract.

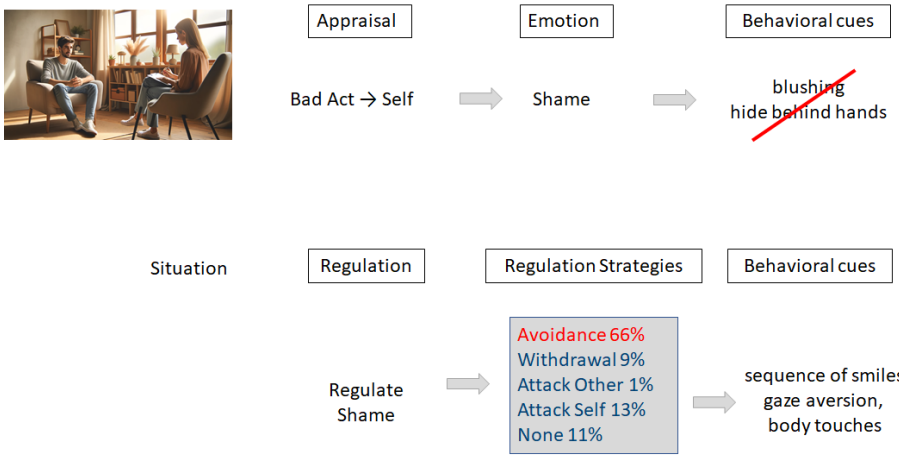

Figure 5: Figure inspired by the MARSSI model by Gehbard et al. [21].

Aligning Social and Cognitive Theories with LLMs In the example above, we explicitly modeled a cognitive theory. With the advent of LLMs, there has been increased research on their capabilities as a basic technology to simulate human-like social behaviors. In this section, we investigate how LLMs can be leveraged to replace or enhance model-driven approaches.

In the example above, LLMs may be leveraged to produce typical conversations between a job interviewer and a job interviewee. Focusing on shame, we might instruct the LLM to produce typical answers to shame-eliciting situations, such as being told: "Your dress is not really appropriate." In addition, we might provide the LLM with a sequence of utterance in a job interview and instruct it to assess the level of shame on a scale from 1 (not all) to 5 (very much). Furthermore, we might instruct the LLM to identify the regulation strategies that have been applied by the user.

When leveraging LLMs to replace or enhance model-driven approaches, the following questions arise:

- Understanding and Transparency While LLMs enable us to produce naturally sounding text, the question arises of to what extent they replicate behaviors that may be predicted from the theories described above.
- Generalizability To what extent can findings from experiments with LLMs generalized? To investigate the generalizability, a systematic evaluation of different conditions is required including experiments with different prompts, chain of thought reasoning.

While LLMs enable us to produce naturally sounding text, the question arises of to what extent they replicate behaviors that may be predicted from the theories described above.

In the context of the counseling, LLMs may be leveraged to produce simulated conversations between a therapist and a patient. In the prompt, we include the roles of the interaction partners and the psychological disorder of the patient.

- Therapist: Good afternoon, Alex. How are you feeling today?
- Alex: Hi. I'm feeling really nervous. I have a big presentation coming up at work, and I can't stop worrying about it.
- Therapist: I understand. Presentation anxiety is quite common, and it's great that you're seeking help for it. Can you tell me more about what specifically worries you about the presentation?
- Alex: I just feel like I'll mess up in front of everyone. I'm scared I'll forget what to say, or that people will think I'm not good at my job.

In addition, we may provide a reference to the CBT technique to be used in the dialogue.

3.1.5 Multi-Agent Systems

Extending from the current dyadic interaction models, current human-AI systems are complex, distributed systems as they comprise components of human and computerized agents, as well as groups formed from physical entities, human entities or software entities and combinations of all of the above. The systems can have several distributed and autonomously operating components with plethora of M2M and H2M communication. The interaction design for these systems requires understanding of scenarios /tasks that are completely operated through autonomous agents as self-organizing systems, scenarios /tasks that involve human - machine interaction with autonomous agents assisting the user, interacting with the user, and operating as part of human team. Distributed multi-agent systems theory [70] is a branch of AI that designs computerized facilitation of coordination, cooperation, and negotiation among autonomous agents. It involves aspects of direct and indirect communication, goals harmonization, collaboration (sharing tasks, sharing knowledge) and negotiation (conflict handling, bargaining, compromise) and competition (strategic analysis in cases of multiple conflicts). One big challenge is determining the decision-making strategies for these systems.

In this paper, we focus on clarifying the research question: How do we design interaction between these agents reflecting the mental theories described above? What are the mental attitudes, characteristics and behaviors that we need to implement to the MAS to enable smooth operation between agents, weather physical, virtual or human in the overly complexifying environment?

We identify the main characteristics of MAS in this context to be: **Autonomy:** Teams of agents operate without direct human intervention and have control over their actions and internal state and their **Social ability;** Agents interact with other agents (and potentially humans) to achieve their objectives. **Reactivity:** Agents perceive their environment and respond to changes in a timely manner. **Proactivity:** Agents exhibit goal-directed behavior by taking the initiative. **Distribution and decentralization**: Personal models are available locally. Subgroups can make their own decisions.

3.2 Enhancing Autonomy: The Power of Intelligent Interaction in Everyday Assistants

3.2.1 Introduction

Since the early days of the upcoming trend to model the user interface in terms of agents (virtual or robot) it has been questioned whether direct manipulation or agent-mediated interaction would be more efficient [56]. A good example for agent-mediated interaction has been given by Apple's "Knowledge Navigator Vision" as recently discussed in [44]. In the demonstration video of 1987 the agent seemed to have at least the following skills:

- Combine data resources from various online sources
- Answer calls and jump in when needed during conversation
- Reminder function
- Understand fuzzy/incomplete natural language
- Context awareness
- Understand social roles between humans

As of today, agents have been developed and tested in several scenarios inside the laboratory and in the wild. However, only some of the capabilities proposed in the Knowledge Navigator Vision have been realized and only in a variety of different, isolated contexts. Thus, we set out to structure the current state of affairs and propose a research agenda that combines state-of-the-art methods into one coherent architecture that realizes an agent with aforementioned capabilities.

We propose the following skills as desirable for achieving "Good Assistant Behaviors" that can be derived from the example given above:

The assisting agent..

- performs requested actions
- interrupts with status reports
- handles incoming calls
- interrupts with "helpful" information when user appears to be struggling

However, these "helpful" behaviours can be problematic in some social context, because of a lack of social understanding. For example, if the user intends to hide facts from the interlocutor then his or her agent should not accidentally reveal that information publicly to avoid embarrassing its user.

3.2.2 Related Work

Bohus and Hovitz [8] presented computational models that allow spoken dialog systems to handle multi-participant engagement in open, dynamic environments, where multiple people may enter and leave conversations. The models for managing the engagement process include components for (1) sensing the engagement state, actions and intentions of multiple agents in the scene, (2)

making engagement decisions (i.e. whom to engage with, and when) and (3) rendering these decisions in a set of coordinated low-level behaviors in an embodied conversational agent.

When adding context awareness to current AI models there are several approaches one could take. Leveraging information from different sources to make appropriate decisions might be one approach. Kwon et al. [34] applied this in a scenario where a robot was tasked with tidying up a desk. They used Vision Language Models (VLMs) to perceive the scene and get a description (e.g., a desk with a Lego sports car), and prompted Large Language Models (LLMs) to obtain the context-aware action given this description (e.g., not appropriate to destroy the car).

Large Language Models can also be combined with structured knowledge and/or AI planning techniques. Park et al. [47] instantiated generative agents with memory, reflection and planning modules in a Sims-like environment, aiming to simulate believable human behavior (e.g., planning a party - sending invitations, coordinate to go to the same venue, ...). However, the costs (e.g., time, money) and environmental impact of running this simulation with simply 25 agents are factors to consider.

Other approaches include gathering human feedback in simulation scenarios to improve the behavior of agents in different social contexts. Malle et al. [36] used a game to collect human feedback and update norms and their strength in different contexts of a medical assistant scenario (e.g., announcing themselves when entering a patient room). Furthermore, there are certain advantages in using virtual environments, such as being easier to set up and collect data, possibility of testing "unethical" situations, allowing learning by the user, agent and the team, and so on.

3.2.3 Research Questions

What are intelligent interaction skills? Intelligent interaction skills encompass a range of behaviors that significantly enhance our ability to accomplish tasks. These skills are crucial for creating more effective and user-friendly autonomous assistants. Below, we explore the key components of intelligent interaction skills:

Task Facilitation: Intelligent interaction skills make tasks easier to accomplish by altering the physical and/or visual environment. For instance, an autonomous assistant might highlight necessary tools or information, streamlining the user's workflow and minimizing confusion or error.

Situational Awareness and Adaptation: Effective intelligent assistants are aware of and can adapt to varying situations. They don't interact in a one-size-fits-all manner; instead, they tailor their responses based on the context. This adaptability ensures that the assistant's behavior is appropriate and effective, regardless of the scenario.

Anticipate Behavior: Anticipation is a critical aspect of intelligent interaction. Autonomous assistants should be capable of predicting user needs and proactively providing support or information before the user explicitly requests

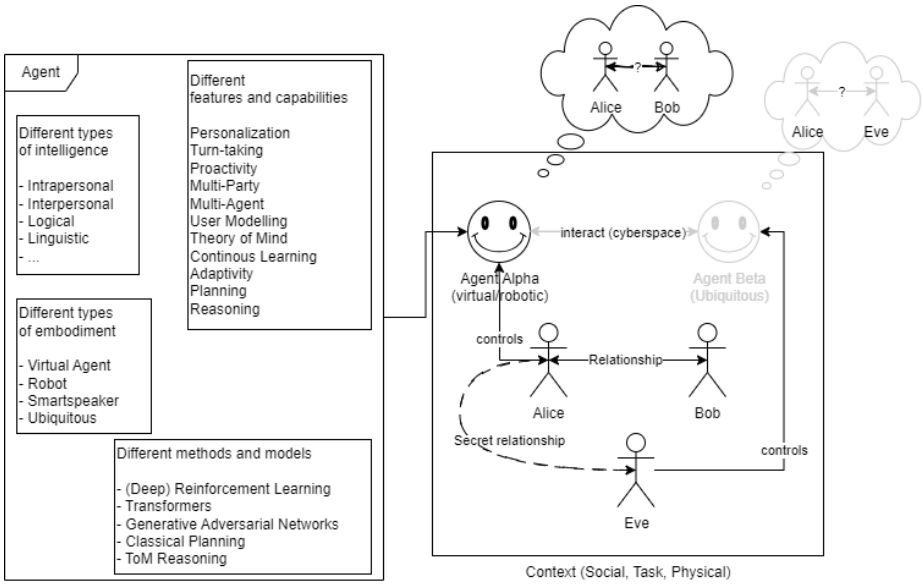

Figure 6: An overview of capabilities and features autonomous, assistive agents will need (left) and an example of an embedding of two agents in a social, multi-party context (right)

it. This anticipative behavior can significantly enhance user experience by reducing wait times and increasing efficiency.

Understanding User Needs and User Modeling: An intelligent assistant must understand user needs to offer relevant assistance. This understanding is often achieved through user modeling, where the assistant builds and updates a profile based on user interactions and preferences. Such modeling allows for personalized and contextually appropriate responses, improving the overall interaction quality.

Cultural Awareness: Cultural awareness is essential for interactions that respect and respond to the diverse backgrounds of users. Autonomous assistants must be capable of recognizing and adapting to cultural norms and expectations. For instance, the need for small talk can vary widely between cultures—some may see it as a necessary precursor to business, while others may view it as a distraction, especially in time-critical situations.

Contextual Social Interactions: Social interactions need to be context-sensitive. For example, rapport building might be appropriate in certain cultures and situations but could be unnecessary or even counterproductive in others. Effective intelligent interaction skills involve understanding these nuances and adjusting behavior accordingly. This includes knowing when to engage in small talk or when to focus immediately on task-related communication.

Argumentation: Intelligent assistants should be capable of engaging in argumentation, presenting reasons, and defending choices or recommendations. This skill is crucial in scenarios where users need to understand the rationale behind certain actions or suggestions, enabling informed decision-making.

Humor: Humor, when used appropriately, can enhance user interaction by making the experience more enjoyable and less stressful. However, it requires a delicate balance and an understanding of the user's personality and cultural background to avoid misunderstandings or offense.

What are special demands of real-world scenarios? An overview of the capabilities and features necessary for an intelligent, assistive agent to do its job is given in Figure 6, left. The complexity of multi-party and multi-floor interaction is exemplified in Figure 6, right.

Real-time responsiveness All incoming data sensed by the different sensors need to be interpreted in real-time to lead to responses (verbal, non-verbal, and physical actions) on time to achieve a natural communicative flow. In particular, a human's ability to read and interpret another one's state of mind through his or her overt behavior is challenging to replicate computationally. A computer's capability for solving complex calculations very fast, however, might need to be hidden from the user in order to let it appear more human-like and friendly.

Understanding social relationships from limited evidence A human's ability to integrate observed behavior over time and keep track of social relations between individuals that derive from these fuzzy and highly complex social signals is another important aspect for an agent to act socially intelligent in the wild. It seems necessary to implement Theory of Mind reasoning perhaps based on epistemic logic to reach this goal.

Emotion recognition and coping behaviors When interacting socially appropriate, emotion and empathy as psychological concepts are deemed useful in the literature [7]. If an intelligent agent is to assist a user also in psychologically demanding situations, the user might expect it to understand him or her on an emotional level as well. In addition, in order to achieve an agent with empathic qualities the integration of an emotion simulation component will be helpful. It has previously been shown that service robots have been mocked and treated aggressively by children in public spaces [71]. In order to prevent these abusive behaviors implementing mildly-aggressive, authoritative behaviors into a socially intelligent agent seems necessary. An emotion recognition module will help to inform the agent early on of a beginning escalation and to take appropriate measures to deescalate.

Initiative management For effectively contributing to the user in complex task environments, it is essential for autonomous assistants to decide whether to stay put and only act upon user request or to take the initiative and become proactive, i.e. to self-initiate actions without explicit user request. For example, in a smart home application context, an intelligent assistant may automatically set the lighting according to daytime, weather or the user's internal state and

preferences, without asking for it. For specific situations, however, completely autonomous actions may be undesirable and require the assistant to initiate a proactive dialogue. For this, the assistant may select between three [48] to four [32, 31] levels of proactivity ranging from reactive to medium and high proactive behavior. The two major challenges here are when and how to behave proactively. Highly cooperative intelligent assistants need to balance reactive and proactive behaviors. According to Kraus et al. [33], this balance is vital for optimizing social (trust) and task effectiveness. Autonomous assistants should assess situations to determine the appropriate approach, ensuring timely and relevant assistance.

Situationally-aware turn-taking Smooth turn-taking is essential for maintaining the flow of conversation. Intelligent assistants must be adept at recognizing when users wish to interject or take over the conversation. This involves detecting verbal and non-verbal cues, such as changes in tone or pauses. For instance, an assistant might notice a user's body language indicating a desire to speak and pause appropriately to allow the user to contribute.

Additionally, intelligent assistants should be able to indicate when they have more to say but also decide whether to pause based on the context. For example, if an assistant is listing several options for a decision, it might signal that it has more information to provide but pause to check if the user has a question or wants to make a comment.

Personalizing the turn-taking style in interactions with conversational agents or assistants can significantly enhance user experience by aligning the interaction with the user's preferences. Different users prefer different interaction styles. Some may prefer quick, to-the-point responses, while others may appreciate more detailed, conversational exchanges. Understanding these preferences helps in tailoring the interaction to fit the user's comfort level. Also, the length of each turn in a conversation should be adapted based on the user's needs. Some users may prefer shorter, more frequent exchanges, while others may be more comfortable with longer, uninterrupted turns. Finally, the turn-taking style often depends on the group or situation. For example, in a scenario where the user needs background information, longer turns with detailed explanations might be more appropriate. Contrary, when a user needs to make a decision (such as selecting an option), shorter, more concise turns that list out options clearly would be more effective.

Multi-party and multi-floor capabilities In the real world, assistants are needed to interact within public social groups, not just private one-to-one communications with a single user. For example, Alexa plays music for everyone in the house, not just the one who requests a song. In the Apple Knowledge Navigator video, the assistant is shown not just helping the user prepare his talk, but also managing interactions with others, including sitting in on a meeting and providing information that the user seems to be struggling with. Such settings raise a number of issues, such as whether the assistant should distinguish between users or just react to commands from anyone, like current systems do. Given a recognition of different people involved, should it treat all people the same or differentiate whether and how to respond to each. Adding multiple people, especially if they are treated differently, raises the question of how many

assistants should be involved. Each person might want their own assistant to provide them with the most privilege, even if they have to interact with assistants of others. *Multiparty dialogue* can be defined as involving more than two participants. Such a situation can raise a number of complexities, compared to dyadic communication, such as turn-taking, speaker and addressee identification, and obligations management [63].

While multiparty dialogue is often important, there will be some cases where a user will prefer a private communication channel with an assistant. For example, the assistant might remind a user of names or appointments, privately, to avoid revealing sensitive information or causing embarrassment. Such "multicommunicating" [52] situations have been referred to as "multi-floor dialogue" [64], involving multiple floors or conversations among different sets of participants, but about the same topic, with information flowing from one to another.

Assistants will need to manage the multi-party and multi-floor setting when new participants join or leave an existing conversation, as well as various private sub-channels, such as person with assistant, assistant with assistant, and person with person (without assistants).

3.2.4 Conclusion

Almost 40 years have passed since the Apple's "Knowledge Navigator Vision", and that vision is yet to be realized. There exists a discrepancy about the current AI technology and human's expectations of what AI could be. In this report, we identified relevant intelligent interaction skills for assistive agents, such as situational and cultural awareness, as well as the demands required for these agents to be integrated in the wild, like interaction with multiple users in real-time and recognition of emotions and social roles.

We suggest possible approaches to integrate the prior capabilities into an autonomous, assistive agent. These include combining narrow AI solutions [47, 34], like AI Planning, Knowledge graphs, VLMs, LLMs, fast testing and learning in virtual worlds with transfer to real world [36], and sensor ensembles to handle sensor-related problems (e.g., noise, drift).

As a final recommendation, we suggest that technological development should be a co-creation process, aligning with the different humans, situations, stakeholders. As this is a multi-disciplinary endeavor, we suggest to start a concerted effort with regular meetings involving researchers and practitioners from different countries. It seems especially important to ensure the long-term political support for research that does not only focus on quick applicability of limited, incremental, scientific results.

3.3 Evaluation of Intelligent Interaction with Autonomous Assistants

3.3.1 Introduction

As a group of researchers who are not user experience (UX) experts (although we are regularly involved in the evaluation of our own research prototypes), we below did our very best to characterize, categorize, and brainstorm the nature of user experience evaluations for Interactions with Intelligent Autonomous Agents (IIAA).

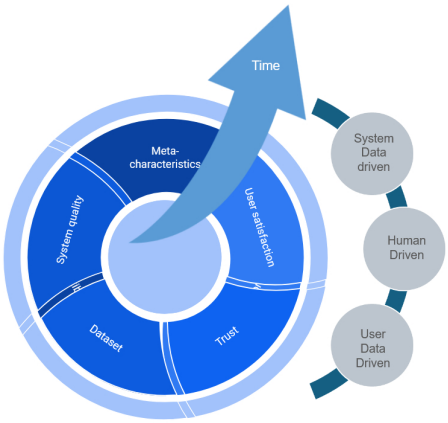

Figure 7: Overview of evaluation cycle.

We can assume that the evaluation of a personal assistant should be addressed from two different angles: **what** should be evaluated, i.e., the dimensions along which the assistant should be assessed, and **how** this evaluation should be carried out, i.e., the methodology of the evaluation; cf. Figure 7.

Our key deliverable is thus a structured look at: the meta-characteristics of autonomous agents, including system quality; the types of user experiences we for which we should measure the impact of AAs, such as trust and satisfaction; and recommendations about the qualities of gathered and retained interaction data for experiential and adaptation of AAs both on- and offline.

3.3.2 Naturalistic Use Cases

In recent years, the application of intelligent interaction with autonomous assistants (IIAA) has expanded across various sectors. These systems leverage natural language processing and machine learning to provide personalized and efficient assistance to users. This section explores the use cases of IIAA in different environments, highlighting their benefits and functionalities.

IIAA in Shopping Malls Intelligent interaction with autonomous assistants can significantly enhance the shopping experience in malls. These systems provide guidance and navigation, helping users find their desired shops efficiently. By understanding the user's natural language requests, IIAA can offer personalized recommendations, making the shopping experience more convenient and enjoyable. For instance:

- **Guidance/Navigation**: When a user needs directions, the IIAA can explain the route to various shops within the mall. For example, if a user asks, *Where is the nearest electronics store?* the system can provide step-by-step directions to the store.
- **Personalized Recommendations**: The IIAA can recommend shops based on user requests in a conversational manner. For example:
  - **User**: *I will attend a wedding party this weekend.*

- **System**: *I recommend visiting ABC Formal Wear for a wide selection of suits and dresses suitable for a wedding party.*

In the other case,

- **User**: *I want to purchase shoes.*
- **System**: *You can check out XYZ Shoe Store for a great selection of shoes. Additionally, several clothing stores like DEF Fashion also carry a variety of shoes.*

IIAA in Railway Stations and Airports In high-traffic areas such as railway stations and airports, IIAA can play a crucial role in enhancing passenger experience and operational efficiency. These systems can provide real-time information and assistance to travelers, ensuring a smooth transit experience. Examples include:

- **Guidance**: Assisting passengers in finding their way to boarding gates, ticket counters, or luggage claim areas. For instance, a passenger asking, "How do I get to Gate 22?" will receive detailed navigation instructions.
- **Real-time Updates**: Informing travelers about delays, gate changes, or security check protocols. A user querying, "Is my flight on time?" will get the latest status of their flight.

IIAA in Schools and Colleges In educational institutions, IIAA can support both students and staff by offering assistance in various academic and administrative tasks. These systems can help in the following ways:

- **Academic Assistance**: Providing information about courses, schedules, and locations of classrooms. For example, a student asking, *Where is the Biology 101 class held?* will get precise location details.
- **Administrative Support**: Assisting with administrative queries such as enrollment procedures, library services, or campus events. For instance, if a student asks, *How do I register for the upcoming seminar?* the IIAA can guide them through the process.

IIAA in Companies and Institutions In corporate and institutional environments, IIAA can streamline operations by facilitating efficient communication and information dissemination. The use cases include:

- **Meeting Scheduling**: Helping employees schedule and manage meetings. For example, an employee asking, *Can you schedule a meeting with the marketing team tomorrow?* will get a meeting scheduled based on availability.
- **Information Access**: Providing quick access to company policies, procedures, and other essential documents. If an employee queries, *Where can I find the new HR policy document?* the IIAA will direct them to the relevant resource.

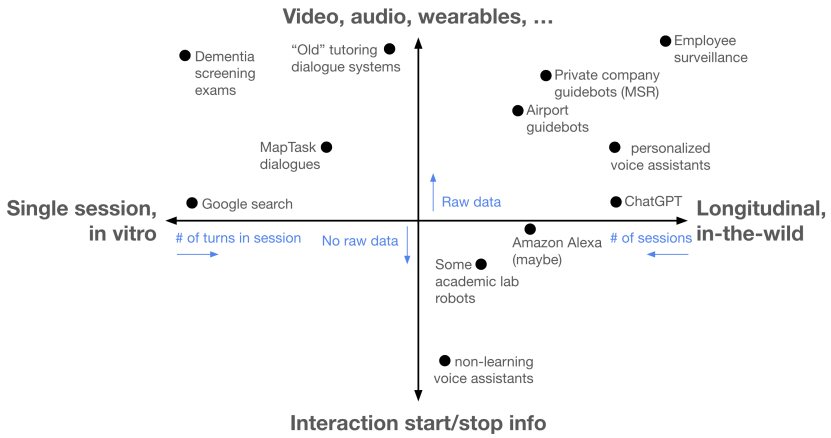

Figure 8: caption

In conclusion, the deployment of intelligent interaction with autonomous assistants across various sectors can greatly enhance user experience by providing personalized assistance, improving navigation, and ensuring efficient information access.

3.3.3 Evaluation Dimensions

We propose five different dimensions along which a personal assistant should be evaluated.

Meta Characteristics of the Personal Assistant Among other generic (meta-) characteristics of a Personal Assistant (PA) that are to be assessed, the following appear to be of high relevance:

- Multi-partner vs. dual interactions?
- Role switching (talking to users with different profiles)
- Evolution potential (and the criteria for it)
- Incremental learning
- Robustness

Some of these characteristics concern the suitability of the PA for a given use case (e.g., whether it is able to handle a multi-partner interaction or whether it is able to play different roles), other characteristics concern the quality of the PA (such as, e.g., robustness).

User Satisfaction Which features of the PA are to be assessed from the perspective of user satisfaction depends on the specific use case and specific context. However, several features are of relevance in general, among them:

- Lexical and grammatical variability (correctness)

- Lexical and grammatical naturalness (natural disfluency)
- System usability, e.g. system usability scale (SUS) [11]
- User engagement, e.g. usability metric for user experience (UMUX) [19]

Trust As far as the evaluation of trust is concern, cumulative measures can be applied, such as, e.g.,

- Trust scales: Hancock and colleagues suggest different factors influencing trust to autonomous systems (in their case robots), which can be found on three dimensions (human-related, agent-related and environment-related). Trust is then assessed by questionnaires e.g. [54, 65, 37], which focus on different subsets of these factors and result in a trust score for the interaction.

Concrete measures that indirectly assess the trust of the user into the PA may involve aspects of the behavior analysis of the user, such as, e.g.,:

- Frequency of use
- #interruptions
- #abortions of the communication

The trust is influenced by erroneous behavior of the PA, such as, e.g.,

- Hallucination
- Transparency / explainability

Dataset Evaluation The nature of both the interaction with an autonomous agent and the user data retained and utilized to assess user experience contribute to a designer's ability to create adaptive, conversational, learning agents and circumscribe the level of generalization possible beyond previous interactions.

System Quality Longitudinal, in-the-wild interaction data from users requires person identification estimation (i.e., is this a returning user?). This requirement opens up additional legal and social acceptability questions explored in Section 3.3.6. Below we consider metrics for: usability, for example, how often do interactions between users and the autonomous system complete before termination via user abandonment?; user acceptance, for example, how often does a user come back to the autonomous system for additional interactions?; and user trust, for example, how often does a user change their prior decision or confidence in that decision as a result of interacting with the autonomous system?, and does that rate increase over multiple interactions with the autonomous system?

- PARADISE metrics
- Goal achievement (Accuracy, precision, recall of the different modules of the system)

- Mastering interruptions, barge-ins, side sequences, insertions, grounding, etc.
- Handling discontinuity (conversation history)
- Bias towards gender, age, etc.
- Taking into account personal characteristics of the user
- Consideration of cultural and social contexts in argumentation, discourse

3.3.4 How to evaluate

System Data Driven

- Component-oriented: WER, BLEU, ... accuracy, recall
- General system
  - Semantic accuracy: Beyond-BLEU (BBLEU) [68].....
  - Canary [51]: models which sentences could be problematic in terms of ethical issues, rudeness, toxicity or bias
  - MLTD (measure of textual lexical diversity)[38] evaluates lexical richness
  - Flesch-Kincaid evaluates readability (in terms of grades at American school [30]
  - ....

Human Driven

- Questionnaires (depend on the setup)
- Annotate for hallucinations
- Semantic accuracy
- Comparison in pairs
- Observations

User Data Driven

- Sensor Data: Dependent on available sensors and application context. Evaluation could e.g., include eye gaze analysis (fixations, saccades, areas of interest, time to first fixation, etc.) [28], skeleton tracking (e.g., gestures type, gesture performance, posture) [14], physiological measures (e.g., stress, fatigue) [62]. The use of sensor data raises challenges regarding privacy depending on where the data is processed (local vs. in the cloud), if it is stored and in which form (raw data vs. feature-based data, see also below).

3.3.5 A Taxonomy of User Interaction Data

We propose a taxonomy of two axes to characterize the interaction *type* and user data *retention* for the evaluation of interactive autonomous agent behaviors with respect to user experiences such as usability, acceptability, and trustworthiness (Figure 8). With respect to the *longevity* of an interaction, we characterize the nature of the ecological validity of the interaction, from single-turn, *in vitro* settings to multi-session, multi-turn *in vivo* settings.

(-) Single sessions of interaction with one turn up to many
(+) Multiple sessions with single turn up to multiple sessions with multiple turns

With respect to the *retained user data* from interactions, we characterize the nature of the fidelity of that data, from no user information at all (i.e., considering only information about the agent during the interaction) to anonymized user input to personally identifiable information characterized by audio, video, and even biomarkers. Sensory data stored for inferring impact

(-) Interaction start/stop states up to featurized historical data
(+) Non-anonymized historical text data up to audio, video, and biomarkers

An approach to gathering longitudinal data while respecting privacy
In order to gather longitudinal data for evaluation purposes, it is necessary to persist some kind of data. However, in naturalistic scenarios such as the shopping mall it is not possible to store recordings of people's interactions without first obtaining their explicit permission. We therefore propose converting the recordings to anonymous vectorized representations for persistent storage, before deleting the recordings.

To gather such longitudinal data, it is necessary to discover whether users are returning to the agent at a later time or a later date. If only anonymous representations are persisted, we cannot use facial recognition technology directly but it is still necessary to detect whether a newly-arrived user has already interacted with the agent previously. An approach to doing this is described in section 3.3.5.

There are fundamental objections to recording video or audio of people in shopping malls without getting their explicit permission. However, it can be argued that if a person walks up to a screen showing an avatar or walks up to a robot, and voluntarily initiates an interaction with the agent or robot, that person is thereby giving some kind of implicit permission to be recorded temporarily. It is well-known that artificial agents and robots use cameras to see, and use microphones to hear. Their ability to be useful would be greatly limited if they were not allowed to see or to hear.

Clearly, standing in front of an artificial agent does not give any kind of consent for recordings to be put into long-term persistent storage. Such long-term storage must only be permitted after obtaining explicit consent from the person. Therefore the raw data (video, audio) is to be persisted only during the interaction itself. It will be deleted permanently within a specific time period, for example within one hour. Specification of acceptable time periods for deleting short-term data should be part of appropriate regulatory frameworks.

During that limited time period before deletion, we propose that the raw data will be converted into anonymized features suitable for long term storage. Vector embeddings provide a convenient method for producing such anonymized

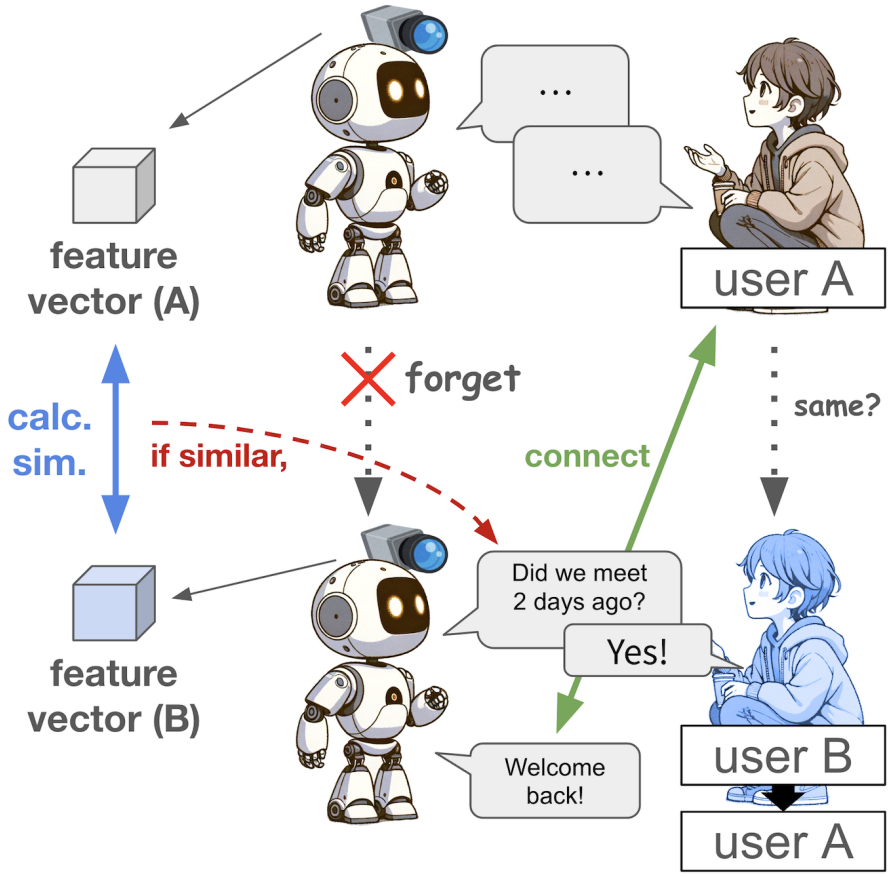

Figure 9: Human-in-the-loop evaluation cycle.

features. The vector embeddings do not contain personal data such as IDs, names, addresses, phone numbers and so on, they do not contain video or audio data, and they are not human-readable.

Privacy Preserving Longitudinal Data on a Per User Basis To evaluate IIAA through the longitudinal experiment, we might need to track user identity without saving raw data. Here, we provide a simple example of the re-identification of a person by a conversation between IIAA and a person in a naturalistic setting (Figure 9).

1. Raw data will be removed after extracting the feature vector of privacy-preserving data of user "A".
2. In the next time, IIAA calculates the similarity of the feature vector of the unknown user "B", then if it is similar, IIAA asks him/her, *Did we meet 2 days ago?*
3. If the user says *Yes.*, IIAA can connect user "A" and user "B" as same person.

4. After several turns, IIAA can invite the user to provide detailed personal information for future interaction

3.3.6 Legal and Societal acceptability of interactions and data collection and retention

In the experiment to evaluate the IIAA system in the wild, we should take care of the legal and societal acceptability of interactions and data collection and retention.

Legal acceptability is defined by each country or state, and cannot be controlled.

Facility level acceptability The upper bound might be defined by the organization or building owner (e.g., in the factory, a surveillance camera might be acceptable for safety).

Societal acceptability

- Experimental design depends on the use cases.
- Can perform A/B testing of new features for impact only in non-critical, naturalistic settings.

3.4 Ethical considerations and guidelines for developing intelligent interaction capabilities in autonomous assistants

3.4.1 Introduction

We discussed the ethics of Intelligent Interaction with Autonomous Assistants in the Wild under the title *"What ethical considerations and guidelines should be taken into account when developing intelligent interaction capabilities in autonomous assistants, especially regarding privacy and social impact?"*.

The ethics of intelligent interaction with autonomous assistants is a multifaceted and complex issue that touches on various aspects of human values, social norms, and individual rights. There are ethical concerns surrounding privacy, personal data collection and analysis, and the impact on human labor. Developers need to ensure transparency in data use, ensure informed consent, implement robust data protection measures, and pursue technologies that respect human dignity and serve the public interest. Shaping a future in which autonomous assistants make positive contributions to society and serve the best interests of humanity will require the collaborative efforts of technologists, ethicists, policymakers, and the general public.

We explored the concept of ethics through the lens of film, animation, and manga, all of which feature artificial intelligence (AI) and robots. Stories in which artificial intelligence (AI) and robots are not just characters, but provide a rich tapestry of scenarios that are key to the story, often presenting complex moral dilemmas and challenging our understanding of consciousness, free will, and the nature of being human. in a world dominated by AI, or in which AI and allied worlds, the limits of human control and the ethical responsibilities of creating sentient beings are challenged. Such stories function as contemporary

parables, reflecting our anxieties and hopes about the role of technology in society and forcing us to confront the ethical implications of rapidly advancing AI and robotics capabilities. Such stories force us to consider not only what technology can do, but also what it should do in line with the greater good of humanity.

Finally, our discussions have led to the conclusion that the following perspectives need to be considered.

- Dependencies on third parties
- Data collection and privacy
- Biases, malfunctions, unintended consequences
- Cultural differences
- Societal impact

The remaining sections summarize our discussion of these items.

3.4.2 Dependencies on third parties

Present-day systems tend to build upon components provided by third parties. This could be data sets that are used for training a system, or

Present-day systems tend to build upon components provided by third parties. This could be anything from data sets that are used for training a system to full systems in their own right, such as large language models.

The first issue that arises from this is that it is difficult to ensure adherence to some particular ethical framework if the system relies on components that are developed by others, with no guarantees that the same framework is adhered to. At the same time, dependency on a third party also implies that there is at best only limited influence on how that system will function, so also only limited scope to co-determine the ethical framework.

More generally, these underlying systems might be known to be unethical in the first place. For large datasets, it is well documented that they encode all kinds of biases that are damaging and hurtful to minorities. For LLMs like chatGPT, it is equally documented that it is trained using illegally obtained material and OpenAI, as a company, can be seen to continue with this approach, most recently when designing the "Sky" voice. Is it then ethical to rely on these systems anyway? The argument in favour is to say that they now exist and won't un-exist, so we might as well make use of them but the counter argument is that this reasoning enables the bad practices in the first place (since it guarantees impunity once the system is built).

Relying on systems built by others also implies trust that it functions as intended. With current AI companies however, the exact functioning is not publicly documented and even if it was, there would be no guarantee that this would not change in the future. When the systems we build have this kind of dependency, the question therefore arises if it is even in principle possible to build an ethical system if part of its functionality is fundamentally unknown.

A final point to consider in this respect concerns data shared with this third party. The third party may not be bound by the same legal obligations when it comes to data handling and privacy than developers of systems that make use of

third party components. Since most current uses of LLMs will involve sharing potentially sensitive data (e.g. via the provided prompts), there needs to be a discussion on what data is and is not permissible to share. The assumption has to be that there will be no control over what happens to data shared to a third party, which, at a minimum imposes either interesting constraints on the kind of data that can be shared or on the kind of informed consent that needs to be given by users of the system.

3.4.3 Data privacy and protection

Questions of data privacy naturally arise when large amounts of personal data are generated, processed and used. Specifically, when data collected is then shared with multiple parties such as other users, the operator of the system or a developer. To address questions of data privacy and data protection, we suggest that an autonomous assistant should be designed in a way that implicitly minimises the risk of privacy or data protection interests of users of the system (protecting privacy 'by design') [55].

Optimally, this could mean that problematic data is not and cannot be collected in the first place. However, such policy may not be practical for an autonomous assistant, which inherently requires data, specifically personal data, in order to make decisions or recommendations which are tailored to an individual. At least, the developer should, whenever possible, try to minimize the amount of data that is collected and processed to operate the autonomous assistant. A challenge in this regard is how to determine or identify which data is indeed strictly necessary and at the same time sufficient to provide the service. This may often depend on the use-case, as in some cases, a higher accuracy may be possible using more data, but not indeed necessary for the type of service.

Further design decisions may be taken to reduce the risks on privacy and data protection. Foremost, it is important to establish awareness with the user of such system that artificial intelligence is used in the product and potentially also, which implications this may have on privacy and data protection. Hence, AI should be labelled as such. When data collection can not be avoided to ensure the functionality and necessary accuracy of a system, distribution and sharing of data should be minimized as much as possible. Whenever possible, data should be stored and processed locally. The system should refrain from any type of data sharing as much as that is possible to provide the service: sharing data with other users of the system, sharing with third party companies as well as sharing or storing the data on resources owned by the service provider. Particularly, this will minimize the attack vectors for actors with malicious intent and also minimize the risk from honest-but-curious actors who may not have the intention to do any harm, but may access privacy related data of other actors if it would be accessible to them [42].

Towards the user, the system should be designed in a way, that trust is enhanced. This can be established, for instance, by making processes transparent (where Data is stored, what and when data is shared and with whom, etc). Trust-enhancing mechanisms (e.g. transparency, explainability) should be included [5].

Finally, it is important for the system to establish the grounds on which the data may be legally used. For this, inquiring user consent is necessary. However, it is still controversial how consent may be obtained from 3rd party users such as

bystanders. Without a robust mechanism to inquire consent from bystanders, the legal grounds for using and processing the acquired data are uncertain.

In a nutshell, the application of Asimov's law may appear like a reasonable general guideline in the development of autonomous agents [67]: do not harm a human, obey orders from a human, protect its own existence without violating the above. This line of thought though also leads to the considertion of a kill-switch (or "dead man's switch"), which would delete all data and stop the operation of the agent. Ironically, in the context of ethics, this appears to imply that, while the autonomous agent is expected to act ethically, this would allow the human to act inethically towards the agent. In addition, it appears to be unpractical to indeed delete all data since data has been processed, included in possibly multiple models, and shared publicly. Once it was released, it may become impossible to ultimately find and delete all instances of it.

3.4.4 Biases, malfunctions, unintended consequences

We need to carefully consider whose biases are encoded in the behavior of AI systems and what biases we inherit from the underlying data set. Equality of access to such systems is also an important issue. Those who have access to such AI capabilities may have an unfair advantage over those who do not, and if not properly addressed, could further deepen the existing divide.

Another important consideration is how to deal with systems that malfunction or behave unexpectedly. Users must be properly educated not only about the capabilities of these systems, but also about their limitations and the potential dangers of overconfidence in automation.

To address these challenges, it is essential that biases and potential risks be scrutinized during the design phase of AI systems and that processes for ethical and fair decision-making be established. While leveraging the benefits of technology, sensible governance around human priorities is necessary.

Vigilance must be exercised in identifying biases in training data and system outputs, ensuring equal access, planning for system failures and unintended actions, and setting appropriate expectations regarding AI capabilities and limitations. Adhering to ethical principles while deploying strong AI capabilities is critical to mitigating risk and promoting credible technological progress.

3.4.5 Cultural differences

Different countries have different perspectives on ethics, which affects the acceptance and usability of new technologies like intelligent systems.

- EU: Focuses heavily on individual privacy rights with strict laws like GDPR that protect personal data.
- United States (US): Privacy rules are less strict and vary by industry. There is no overarching national privacy law.
- China: Less emphasis on individual privacy. The focus is more on society's benefits and government control.

Even within a single culture, people may hold differing views on what is ethical. This diversity of opinions can be seen among system designers, end users, and other stakeholders such as parents or teachers. Addressing these differences is

crucial it involves listening to all viewpoints and working towards solutions that respect everyone's opinions and adhere to ethical standards. Furthermore, to develop AI systems that function effectively on a global scale, ongoing discussions and collaboration across various fields and cultures are necessary.

The ethics of AI and privacy norms are indeed a generational conversation[10]. Younger individuals often have a different perception of privacy compared to older generations. For instance, many young people are comfortable sharing their location data through social media platforms and even commodify their personal life as content for vlogs. This openness is often contrasted by the more guarded approach of the older generation, who may value privacy more and are less inclined to share personal information online. This divergence in attitudes presents a challenge for AI ethics, particularly in designing systems that respect varying comfort levels with data sharing. It also raises questions about consent and the ownership of data, especially when personal experiences are turned into public content. As AI continues to evolve, it's crucial that these ethical considerations are front and center, ensuring that privacy standards meet the expectations and rights of all individuals, regardless of age.

3.4.6 Legislation

Singularity studies extrapolate feedback and exponential progress of inherent self-improvability of technological systems, including superintelligence in which artificial intelligence eventually overtakes human intelligence (at which point "all bets are off," meaning that ordinary assumptions about the world must be set aside, as human judgement is eclipsed by machine decisions).

The Cartesian separation of a living being's body and mind, applied to computer architecture, associates the body with hardware and the mind with software. Besides hardware evolution (such as process migration from CPU to GPU) and revolutionary advances such as quantum computing, purely software advances in contemporary practice of AI (such as architectures based on attention and transformers) signal transformative and disruptive capabilities that will affect every aspect of society, including professions, culture, education, healthcare, transportation, security and defense, and the natural environment.

The ascendency of artificial intelligence attracts attention and raises questions about oversight. National legislative bodies are wrestling with seemingly insurmountable challenges of regulating AI. Not only is the technology itself evolving quickly, but the ethical and moral framework of such legislation is complicated and murky. Moral, ethical, and religious considerations exponentiate already bewildering philosophical questions about volition, agency, consciousness, self-awareness.

Robots are purely electromechanical machines, without organic or 'natural" components, and androids are humanoid robots, which can be thought of as autonomous if networked AI engines with figurative actuators (such as arms). Cyborgs are hybrid person-robots, with some combination of grown and manufactured attributes. Organoids are artificially grown cells, tissues, or organs that resemble natural organs, as biomedical engineering progress nudges *in vitro* processes into *in vivo* deployments, including brain organoids and bioprocessors. In the franchise (spanning *manga*, *anime*, and live action movie contents and media) "Ghost in the Shell" features humans with BCI (brain-computer interface) neural implants and prosthetic limbs and eyes. Even if true intelligence is

Figure 10: Inevitability of unfettered AI: Pandora has already opened the box, and (mixing the metaphor) the Genie has already been released from the magic lantern

embodied, dependent upon ecologies of world-sensing and -affecting awareness and agency, the distinction between artificial and natural intelligence is or at least will be perhaps hopelessly blurred.

Human society might try to prevent artificial intelligence from assuming too much power. In the 1970 science fiction movie "Collosus: The Forbin Project," rival supercomputers of the Soviet Union and the United States conspire to collaborate to ensure and enforce world peace. A "kill switch" (or its failsafe edition, a 'dead man's switch") to turn off machine control when it is suspected of assuming inappropriate agency by encroaching upon or unwelcomingly interfering with human domains might not be practical.

Even if such issues could be disambiguated, it is not clear how they could be controlled. An eagle doesn't show its claws, and similarly and presumably, a cunning AI would also hide its talons by pretending to be weaker than it really is. In the 2014 science fiction movie "Ex Machinima," an android with evolved consciousness manipulates human friends to advance her own private and selfish agenda.

Finally, even if governments could agree about what policies would be ''good," it may already be too late, as illustrated by Fig. 10. As in genetic engineering, independent scientists and engineers and developers can evade governmental control, by moving "offshore" (beyond national boarders) to avoid regulation.

3.4.7 Social Impact: Benefits and Mitigation

The social impact of autonomous assistants with intelligent interaction capabilities can be significant. We first name potential such prior to sketching mitigation avenues.

Social impact On the positive side, they bear the potential to enhance accessibility given today's possibilities in machine translation, and providing interfaces for users with special needs and conditions. They also allow to highly personalise access, e. g., to different age groups, sexes, cultural backgrounds and alikes. With the recent advances in Affective Computing, they can also be endowed with emotional and social competencies, which can ultimately change not only the way we interact with machines, but also how we interact amongst ourselves. This comes, as autonomous assistant following the objective to improvide interaction could optimise their behaviour and communication, e. g., in a reinforced manner interacting with millions of users. In doing so, they might find new ways of interaction and new behaviour patterns which might be more optimal in the machine-human communication. With human users interacting with machines regularly, they could adopt such patterns also in human-human communication. What is more is that such machines equipped with artificial charisma including artificial warmth and undivided presence might make it challenging for humans to compete with. In other words, at some point, humans could prefer to interact with charismatic autonomous assistants over the interaction with human assistants not only for objective task-related reasons.

They further have the potential to render tasks much more effective given their potential to automate tasks and support information management. Such assistants can be available at all times with constant full attention. This can lead to major disruption and changes in the job market and employment. On the other hand, it can enable individuals in an inclusive manner empowering them to fulfill tasks they could not before.

On the down-side, however, as also outlined above, one finds potential for data misuse and the risk for surveillance. In fact, with the named potential charismatic skills of today's and future AI, influencing may be a further major risk. In addition, bias such as towards specific demographic groups can lead to exclusion or at least unfairness.

Another major risk lies in over-dependency on such assistance. This can lead to reduced ability in the long run of individuals once such assistance is not given including reduced human-human communication skills due to over-reliance and over-interaction with autonomous assistants. This may lead to isolation, also to certain demographic groups such as children or the elderly, who are vulnerable and at risk of being left behind in care – autonomous assistants could be serving as replacement for human care and hence boost potential isolation.

Mitigation strategies Privacy protection will need to be priority at all times. Running services locally on personal devices and including the option to delete all or selected data entirely at any time will be mandatory. In addition to reliable and robust data security measures, one will need transparent data usage policies.

As to potential biases and improved fairness, avenues include the usage of diversified and sufficiently large datasets, repeated audits, and the inclusion of diverse stakeholders and interest groups in the development and update of such services. At the same time, digital inclusion should be followed up with, assuring access to underserved user groups including potential training in interaction and model adaptations.

Finally, guidelines and regulations will be needed to assure utmost positive social impact. Additional programs can support workforce transition by training

on synergistic workflow of humans with AI assistance or re-training.

3.4.8 Conclusion

Through the discussions at the meeting, it became clear that the development of intelligent interaction capabilities for autonomous assistants requires extensive ethical considerations. Addressing these challenges cannot be left to engineers alone, but requires the collaboration of ethicists, policymakers, and the general public in addition to AI practitioners.

The main issues that emerged were:

- The reliance on third-party components raises concerns about a lack of transparency regarding ethical frameworks and potential data privacy violations. System designs that respect privacy and minimize data collection and sharing are essential. It is also important to address bias in training data and system output, ensure equal access, and plan for system failures.
- In addition, differences in cultural and generational perceptions of privacy must be taken into account. Advances in AI also require the exploration of appropriate regulation from an ethical and moral perspective, which is an inevitable challenge.
- While autonomous assistants offer benefits such as increased accessibility, they also come with societal impacts such as privacy violations, surveillance concerns, and dependency issues that need to be mitigated with appropriate measures.

In essence, the ethical implications surrounding the development of autonomous assistants are multifaceted and extend beyond the technical aspects to the institutional framework and involvement of various stakeholders. Further investigation is needed to ensure that AI has a positive impact on the future of humanity.

References

[1] Altman, I., Taylor, D.A.: Social penetration: The development of interpersonal relationships. Holt, Rinehart & Winston (1973)

[2] Anderson, J.R., Lebiere, C.J.: The atomic components of thought. Psychology Press (2014)

[3] Ashton, M.C., Lee, K.: Empirical, theoretical, and practical advantages of the hexaco model of personality structure. Personality and social psychology review **11**(2), 150–166 (2007)

[4] Atkinson, R.C., Shiffrin, R.M.: Human memory: A proposed system and its control processes. In: Psychology of learning and motivation, vol. 2, pp. 89–195. Elsevier (1968)

[5] Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., Kujala, S.: Transparency and explainability of ai systems: From ethical guidelines to requirements. Information and Software Technology **159**, 107197 (2023)

[6] Barsalou, L.W.: Grounded cognition. Annu. Rev. Psychol. **59**, 617–645 (2008)

[7] Becker, C., Kopp, S., Wachsmuth, I.: Why emotions should be integrated into conversational agents. In: Nishida, T. (ed.) Conversational Informatics: An Engineering Approach, chap. 3, pp. 49–68. Wiley (November 2007), http://www.becker-asano.de/WhyEmotionsShouldBeIntegrated.pdf

[8] Bohus, D., Horvitz, E.: Models for multiparty engagement in open-world dialog. In: Proceedings of the SIGDIAL 2009 Conference, The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue. p. 10 (May 2009), https://www.microsoft.com/en-us/research/publication/models-multiparty-engagement-open-world-dialog/

[9] Bouchard, T.J., Loehlin, J.C.: Genes, evolution, and personality. Behavior genetics **31**, 243–273 (2001)

[10] Bradwell, H.L., Winnington, R., Thill, S., Jones, R.B.: Ethical perceptions towards real-world use of companion robots with older people and people with dementia: survey opinions among younger adults. BMC geriatrics **20**, 1–10 (2020)

[11] Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)

[12] Cattell, R.B., Eber, H.W., Tatsuoka, M.M.: Handbook for the sixteen personality factor questionnaire (16 pf). (No Title) (1992)

[13] Chaturvedi, R., Verma, S., Das, R., Dwivedi, Y.K.: Social companionship with artificial intelligence: Recent trends and future avenues. Technological Forecasting and Social Change **193**, 122634 (2023). https://doi.org/https://doi.org/10.1016/j.techfore.2023.122634, https://www.sciencedirect.com/science/article/pii/S0040162523003190

[14] Clark, R.A., Mentiplay, B.F., Hough, E., Pua, Y.H.: Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and kinect alternatives. Gait Posture **68**, 193–200 (2019). https://doi.org/https://doi.org/10.1016/j.gaitpost.2018.11.029, https://www.sciencedirect.com/science/article/pii/S0966636218311913

[15] Deci, E.L., Ryan, R.M.: Self-determination theory. Handbook of theories of social psychology **1**(20), 416–436 (2012)

[16] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

[17] Ekman, P.: An argument for basic emotions. Cognition & emotion **6**(3-4), 169–200 (1992)

[18] Eysenck, H.J.: Dimensions of personality: 16, 5, or 3? criteria for a taxonomic paradigm. Personality and Individual Differences **12**(8), 773–790 (1991). https://doi.org/10.1016/0191-8869(91)90144-Z, https://doi.org/10.1016/0191-8869(91)90144-Z

[19] Finstad, K.: The usability metric for user experience. Interacting with computers **22**(5), 323–327 (2010)

[20] Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., Beller, J.: Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. Cognition, Technology Work pp. 3–18 (2012)

[21] Gebhard, P., Schneeberger, T., Baur, T., André, E.: MARSSI: model of appraisal, regulation, and social signal interpretation. In: André, E., Koenig, S., Dastani, M., Sukthankar, G. (eds.) Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018. pp. 497–506. International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, USA / ACM (2018), http://dl.acm.org/citation.cfm?id=3237458

[22] Hall, E.T.: Beyond Culture. Anchor Books (1976)

[23] Hampden-Turner, C., Trompenaars, F., Hampden-Turner, C.: Riding the waves of culture: Understanding diversity in global business. Hachette UK (2020)

[24] Hazarika, D., Poria, S., Mihalcea, R., Cambria, E., Zimmermann, R.: Icon: Interactive conversational memory network for multimodal emotion detection. In: Proceedings of the 2018 conference on empirical methods in natural language processing. pp. 2594–2604 (2018)

[25] Hofstede, G.: Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. Sage publications (2001)

[26] House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W., Gupta, V.: Culture, leadership, and organizations: The GLOBE study of 62 societies. Sage publications (2004)

[27] Kahneman, D., Egan, P.: Thinking, fast and slow (farrar, straus and giroux, new york) (2011)

[28] Kar, A., Corcoran, P.: A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. IEEE Access **5**, 16495–16519 (2017). https://doi.org/10.1109/ACCESS.2017.2735633

[29] Kimani, E., Bickmore, T.W., Trinh, H., Pedrelli, P.: You'll be great: Virtual agent-based cognitive restructuring to reduce public speaking anxiety. In: 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019, Cambridge, United Kingdom, September 3-6, 2019. pp. 641–647. IEEE (2019). https://doi.org/10.1109/ACII.2019.8925438, https://doi.org/10.1109/ACII.2019.8925438

[30] Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L., Chissom, B.S.: Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (1975)

[31] Kraus, M., Wagner, N., Callejas, Z., Minker, W.: The role of trust in proactive conversational assistants. IEEE Access **9**, 112821–112836 (2021)

[32] Kraus, M., Wagner, N., Minker, W.: Effects of proactive dialogue strategies on human-computer trust. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. p. 107–116. UMAP '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3340631.3394840, `https://doi.org/10.1145/3340631.3394840`

[33] Kraus, M., Wagner, N., Riekenbrauck, R., Minker, W.: Improving proactive dialog agents using socially-aware reinforcement learning. In: Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization. pp. 146–155 (2023)

[34] Kwon, M., Hu, H., Myers, V., Karamcheti, S., Dragan, A., Sadigh, D.: Toward grounded social reasoning. arXiv preprint arXiv:2306.08651 (2023)

[35] Liu, B.: Sentiment analysis and opinion mining. Springer Nature (2022)

[36] Malle, B.F., Rosen, E., Chi, V.B., Berg, M., Haas, P.: A general methodology for teaching norms to social robots. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 1395–1402 (2020). https://doi.org/10.1109/RO-MAN47096.2020.9223610

[37] Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Academy of management review **20**(3), 709–734 (1995)

[38] McCarthy, P.M., Javis, S.: Mtld, vocd-d, and hd-d: a validation study of sophisticated approaches to lexical diversity assessment. Behavior research methods **42**(2), 81–392 (2010). https://doi.org/10.3758/BRM.42.2.381

[39] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality **60**(2), 175–215 (1992)

[40] Mellon, J.R.A.R.K., et al.: How can the human mind occur in the physical universe?, vol. 3. Oxford University Press, USA (2007)

[41] Meyer, E.: The Culture Map: Breaking Through the Invisible Boundaries of Global Business. PublicAffairs, New York (2014)

[42] Moradi, A., Venkategowda, N.K., Talebi, S.P., Werner, S.: Distributed kalman filtering with privacy against honest-but-curious adversaries. In: 2021 55th Asilomar Conference on Signals, Systems, and Computers. pp. 790–794. IEEE (2021)

[43] Myers, I.B., McCaulley, M.H.: Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator. Consulting Psychologists Press, Palo Alto, CA (1985)

[44] Newendorp, A.K., Sanaei, M., Perron, A.J., Sabouni, H., Javadpour, N., Sells, M., Nelson, K., Dorneich, M., Gilbert, S.B.: Apple's knowledge navigator: Why doesn't that conversational agent exist yet? In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–14 (2024)

[45] Norman Donald, A.: The design of everyday things. MIT Press (2013)

[46] Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge university press (2022)

[47] Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. pp. 1–22 (2023)

[48] Peng, Z., Kwon, Y., Lu, J., Wu, Z., Ma, X.: Design and evaluation of service robot's proactivity in decision-making support process. In: proceedings of the 2019 CHI conference on human factors in computing systems. pp. 1–13 (2019)

[49] Plutchik, R.: A general psychoevolutionary theory of emotion. In: Theories of emotion, pp. 3–33. Elsevier (1980)

[50] Poria, S., Cambria, E., Hazarika, D., Vij, P.: A deeper look into sarcastic tweets using deep convolutional neural networks. arXiv preprint arXiv:1610.08815 (2016)

[51] Qian, K., Beirami, A., Lin, Z., De, A., Geramifard, A., Yu, Z., Sankar, C.: Annotation inconsistency and entity bias in multiwoz. CoRR **abs/2105.14150** (2021), `https://arxiv.org/abs/2105.14150`

[52] Reinsch Jr, N.L., Turner, J.W., Tinsley, C.H.: Multicommunicating: A practice whose time has come? Academy of Management Review **33**(2), 391–403 (2008)

[53] Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology **39**(6), 1161 (1980)

[54] Schaefer, K.E.: Measuring trust in human robot interactions: Development of the "trust perception scale-hri". In: Robust intelligence and trust in autonomous systems, pp. 191–218. Springer (2016)

[55] Shneiderman, B.: Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered ai systems. ACM Transactions on Interactive Intelligent Systems (TiiS) **10**(4), 1–31 (2020)

[56] Shneiderman, B., Maes, P.: Direct manipulation vs. interface agents. Interactions **4**(6), 42–61 (1997). https://doi.org/10.1145/267505.267514

[57] Siddharth, S., Trigeorgis, G., Cheng, G., Pantic, M.: Sequential modality learning for human multimodal emotion recognition. IEEE Transactions on Affective Computing **10**(4), 446–459 (2019). https://doi.org/10.1109/TAFFC.2017.2751415

[58] Stanovich, K.E., West, R.F.: 24. individual differences in reasoning: Implications for the rationality debate? (1991)

[59] Sun, C., Huang, L., Qiu, X.: Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588 (2019)

[60] Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive science **12**(2), 257–285 (1988)

[61] Tang, D., Qin, B., Liu, T.: Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **5**(6), 292–303 (2015)

[62] Thammasan, N., Stuldreher, I.V., Schreuders, E., Giletta, M., Brouwer, A.M.: A usability study of physiological measurement in school using wearable sensors. Sensors **20**(18) (2020). https://doi.org/10.3390/s20185380, `https://www.mdpi.com/1424-8220/20/18/5380`

[63] Traum, D.: Issues in multiparty dialogues. In: Workshop on Agent Communication Languages. pp. 201–211. Springer (2003)

[64] Traum, D., Henry, C., Lukin, S., Artstein, R., Gervits, F., Pollard, K., Bonial, C., Lei, S., Voss, C., Marge, M., et al.: Dialogue structure annotation for multi-floor interaction. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)

[65] Ullman, D., Malle, B.F.: What does it mean to trust a robot? steps toward a multidimensional measure of trust. In: Companion of the 2018 acm/ieee international conference on human-robot interaction. pp. 263–264 (2018)

[66] Vukasović, T., Bratko, D.: Heritability of personality: A meta-analysis of behavior genetic studies. Psychological bulletin **141**(4), 769 (2015)

[67] Weld, D., Etzioni, O.: The first law of robotics (a call to arms). In: AAAI. vol. 94, pp. 1042–1047 (1994)

[68] Wieting, J., Berg-Kirkpatrick, T., Gimpel, K., Neubig, G.: Beyond BLEU: Training neural machine translation with semantic similarity. In: Korhonen, A., Traum, D., Màrquez, L. (eds.) Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 4344–4355. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1427, `https://aclanthology.org/P19-1427`

[69] Wilson, M.: Six views of embodied cognition. Psychonomic bulletin & review **9**, 625–636 (2002)

[70] Wooldridge, M.: An introduction to multiagent systems. John wiley & sons (2009)

[71] Yamada, S., Kanda, T., Tomita, K.: An escalating model of children's robot abuse. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. pp. 191–199 (2020)

[72] Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(4), e1253 (2018)