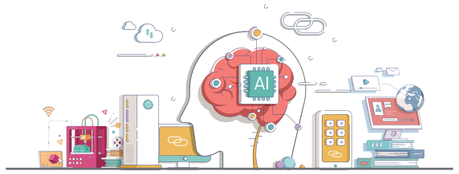


Artificial Intelligence and the Digital Age

Implications for the Future of Society

KAST 한국과학기술한림원 The Korean Academy of Science and Technology

42, Dolma-ro(7-1, Gumi-Dong), Bundang-gu, Seongnam-si, Gyeonggi-do, 13630, Korea
Tel : +82 31.726.7900 Fax : +82 31.726.7909 E-mail : dhkim@kast.or.kr

※ 본 사업은 과학기술진흥기금 및 복권기금으로 지원되고 있습니다.

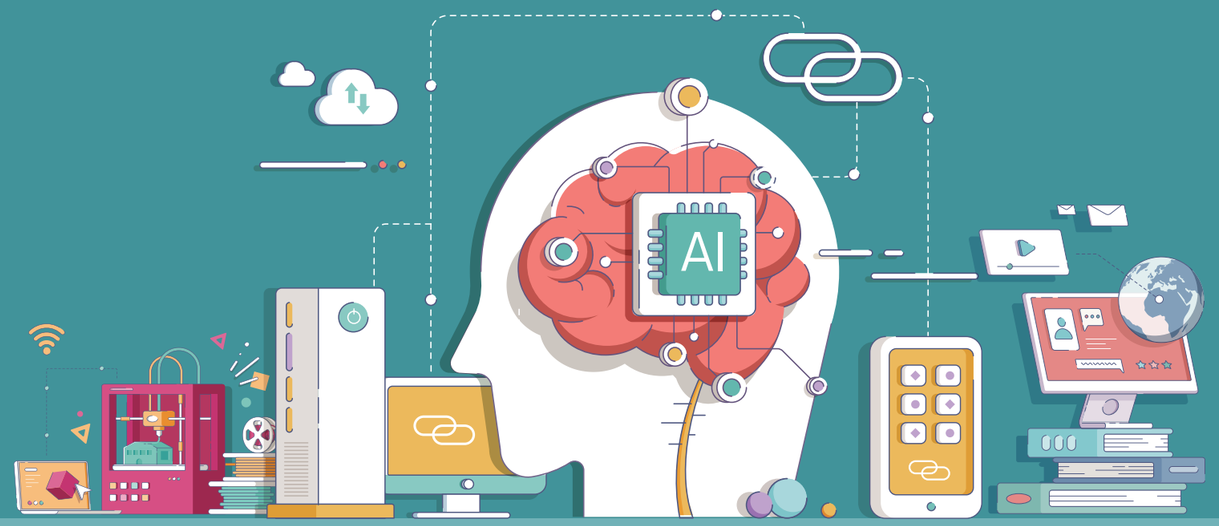
Artificial Intelligence and the Digital Age Implications for the Future of Society

Artificial Intelligence and the Digital Age

Implications for the Future of Society

제6회 한국·독일한림원 공동심포지엄

29-30 September, 2022
The Plaza Hotel Seoul, Ruby Hall



KAST 한국과학기술한림원 The Korean Academy of Science and Technology

Leopoldina Nationale Akademie der Wissenschaften

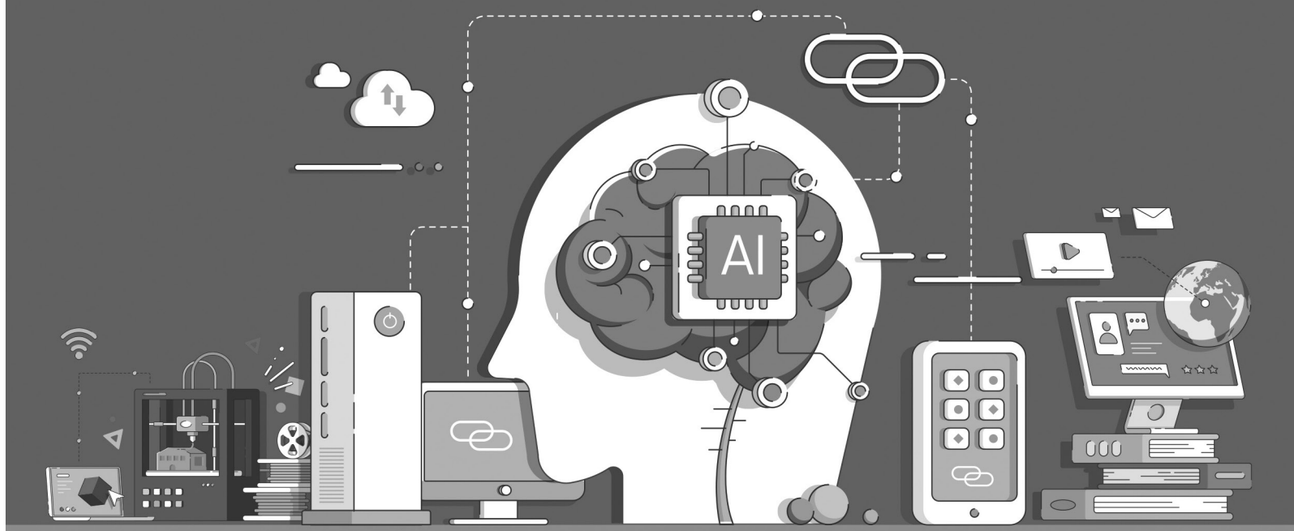
Artificial Intelligence and the Digital Age

Implications for the Future of Society

제6회 한국·독일한림원 공동심포지엄

29-30 September, 2022

The Plaza Hotel Seoul, Ruby Hall



KAST 한국과학기술한림원
The Korean Academy of Science and Technology

Leopoldina
Nationale Akademie
der Wissenschaften

Program

September 29th, 2022 (Thursday)	
09:30~10:00	**Registration**
Commemoration of the 10th Anniversary of KAST and Leopoldina	
10:00~10:15	**Commemorative Address** **Ook Joon Yoo,** President, Korean Academy of Science and Technology **Ulla Bonas,** Vice President, German National Academy of Sciences Leopoldina **Congratulatory Remark** **Michael Reiffenstuel,** German Embassy to the Republic of Korea
10:15~10:25	**MoU Signing Ceremony & Group Photo**
Opening Ceremony	
10:25~10:30	**Opening Remarks from Organizers** **Seong-Whan Lee,** Professor, Korea University **Alexander Waibel,** Professor, Karlsruhe Institute of Technology & Carnegie Mellon University
SESSION 1:	**Current Trends in A.I. Research**
10:30~10:55	**Emerging Trends in Artificial Intelligence: Transparent, Conversational and Applicable AI** **Seong-Whan Lee** (Korea University)
10:55~11:20	**Intelligent Systems and a Language Transparent World** **Alexander Waibel** (Karlsruhe Institute of Technology and Carnegie Mellon University)
11:20~11:45	**Explainable Artificial Intelligence:** **Can we now understand internal mechanisms of deep neural networks?** **Jaesik Choi** (KAIST)
11:45~13:15	**Lunch**
13:15~13:40	**Socially-Interactive Artificial Intelligence:** **Perception, Synthesis and Learning of Human-Like Behaviors** **Elisabeth André** (Augsburg University)
13:40~14:05	**Conversation Modeling for Computational Social Science** **Alice Haeyun Oh** (KAIST)
14:05~14:30	**Discussion**
14:30~15:00	**Tea Break**

SESSION 2:	**The Crucial Role of Data**
15:00~15:25	**Adversarial Learning to Mitigate Algorithmic Bias of AI Systems Caused by Data** **Sung-Bae Cho** (Yonsei University)
15:25~15:50	**Self-Learning Based Defect Reduction for Dialogue Systems** **Eunah Cho** (Amazon)
15:50~16:15	**On the Role of Public Biomedical Data for Personalized Medicine** **Hyunjung Shin** (Ajou University)
16:15~16:40	**Discussion**
16:40~17:00	**Summary Discussion of Conference Day 1 and Outlook on Conference Day 2**

September 30th, 2022 (Friday)	
09:30~10:00	**Registration**
SESSION 3:	**From Application to Acceptance**
10:00~10:25	**AI Methods for Predicting Toxicity of Compounds in Chemical and Genetic Space** **Sun Kim** (Seoul University)
10:25~10:50	**Autonomy and Trust in Self-improving AI Systems** **Byoung-Tak Zhang** (Seoul National University)
10:50~11:15	**How to Create Successful Technology Disruption** **Alexander Waibel** (Karlsruhe Institute of Technology and Carnegie Mellon University)
11:15~12:30	**Lunch**
12:30~12:55	**Fair AI** **Chang Dong Yoo** (KAIST)
12:55~13:20	**Constructivist Psychology and the Digital Age** **Reinhold Kliegl** (University of Potsdam)
13:20~13:45	**AI for Social Impact: Poverty and Disaster Mapping from the Sky** **Meeyoung Cha** (KAIST)
13:45~14:10	**Interfaces of AI and Democracy** **Jeanette Hofmann** (Freie Universität Berlin)
14:10~14:30	**Summary Discussion and Closing Remarks**

Commemoration of the 10th Anniversary of KAST and Leopoldina

Ook Joon Yoo
President
The Korean Academy of Science and Technology (KAST)

ojyoo@kaist.ac.kr

Prof. Ook Joon Yoo served as the Director of BioMedical Research Center and professor of Graduate School of Medical Science and Engineering at KAIST. He also performs an important role in the National Science and Technology council, as a member.

He began his research and teaching at the Department of Biological Science of KAIST from 1982 and he also was the Chairman of the Department. He served as the Director of Molecular Medicine Research Group supported by Ministry of Science & Technology. Besides, He was the Chairman of the Korean Bioscience and Biotechnology Association, and the President of the Korean Society for Biochemistry and Molecular Biology. Also, he served as the Executive-Vice President of the KAST during 2016-2019.

With his distinguished achievements as a molecular biologist, he was received diverse awards including the Order of Science and Technology Merit from Korea Government in 2015, Grand Research Award from KAIST in 2002, Research & Development Award from KAIST in 1996, and Honoring Award (Excellent Inventive Prize) from Daejeon City in 1999.

He received his B.S. from the Seoul National University, and his Ph.D. in Biochemistry and Molecular Biology from the University of Chicago. His research interest covers genomics-based drug target discovery & validation, studies on genetic disorders.

He published more than 140 articles mostly through international journals in his research field. He also is the author of the book, BioMedical Research, which has been one of best and steady sellers for last 25years.

Currently, he is the President of the Korean Academy of Science and Technology (KAST).

Ulla Bonas
Vice President, German National Academy of Sciences Leopoldina

Ulla.bonas@genetik.uni-halle.de
https://www.biologie.uni-halle.de/institutsbereich_genetik/plant_genetics/prof._dr._ulla_bonas/

Professor Dr. Ulla Bonas studied Biology at the University of Cologne from 1974 to 1980. In 1984, she completed her PhD thesis in Genetics at the University of Cologne with a dissertation entitled In-vitro cloning of a transposable element in the chalcone synthase gene of Antirrhinum majus. From 1985 to 1987, she worked as a postdoc at the University of California, Berkeley, supported by fellowships from DAAD, DFG and the Max-Planck society. From 1988 to 1993, she was the Leader of an Independent Research Group at the Institute for Gene Biology Research in Berlin. In 1992, she habilitated in Genetics at the Free University of Berlin with a thesis on Molecular genetic analysis of the interaction between Xanthomonas campestris pv. vesicatoria and the plant. From 1993 to 1998, Ulla Bonas was a Groupleader (Directeur de Recherche; permanent) at the CNRS Insitute of Plant Sciences in Gif-sur-Yvette, France. From 1998 until 2021, she was a Full Professor of Genetics at the Martin Luther University Halle-Wittenberg, Halle, Germany. Her research was focused on the genetic and molecular analysis of pathogenicity and plant disease resistance in the Xanthomonas/plant interaction.

Ulla Bonas is a member of the European Molecular Biology Organization (EMBO) and in 2011 received the prestigious Gottfried Wilhelm Leibniz Price of the German Research Foundation (DFG). Since 2008, she is a member of the German National Academy of Sciences Leopoldina, and since 2015, she is vice-president of the Leopoldina.


© German Embassy Seoul

Michael Reiffenstuel
Ambassador
German Embassy to the Republic of Korea

Michael Reiffenstuel is German Ambassador to the Republic of Korea since 2020. He joined the Federal Foreign Office in 1993. His professional career took him to Sarajevo as Deputy Head of Mission at the German Embassy, to London as Foreign Policy Officer and to New Delhi as Press Officer. Most recently, Michael Reiffenstuel was Consul General in Istanbul.

From 2005 to 2009, he initially headed the Team Deutschlandbild im Ausland (German Image Abroad) and then served as Deputy Head of Division responsible for Strategy and Planning Foreign Communications. From 2009 to summer 2012, he worked first as head of the German Information Centre Arab World and later in the foreign policy section at the Embassy in Cairo. From July 2012 to July 2015, Michael Reiffenstuel was Head of Unit 312, responsible for Transformation in the Arab World and Bilateral and EU Relations with the Maghreb States. He then served as Commissioner for Foreign Cultural Policy and Deputy Director-General of the Division for Culture and Communication until July 2018. His activities at the Federal Foreign Office in Berlin also included international legal issues and non-proliferation of weapons of mass destruction.

Michael Reiffenstuel studied law at the Ludwig Maximilians University in Munich and obtained a Master of Law (LLM) at the University of Georgia/Athens (USA) in 1993.

Memo

The 6th KAST-Leopoldina Bilateral Symposium

Artificial Intelligence and the Digital Age - Implications for the Future of Society

SESSION 1

CURRENT TRENDS IN A.I. RESEARCH

Co-chair

Sung-Bae Cho (Yonsei University)

Eunah Cho (Amazon)

Speakers

Towards Explainable AI with the Relative Contributions in Deep Neural Networks

Seong-Whan Lee (Korea University)

Intelligent Systems and a Language Transparent World

Alexander Waibel (Karlsruhe Institute of Technology and Carnegie Mellon University)

Explainable Artificial Intelligence:
Can we now understand internal mechanisms of deep neural networks?

Jaesik Choi (KAIST)

Socially-Interactive Artificial Intelligence:
Perception, Synthesis and Learning of Human-Like Behaviors

Elisabeth André (Augsburg University)

Conversation Modeling for Computational Social Science

Alice Haeyun Oh (KAIST)

Emerging Trends in Artificial Intelligence: Transparent, Conversational and Applicable AI

Seong-Whan Lee
Professor
Department of Artificial Intelligence, Korea University

sw.lee@korea.ac.kr

Seong-Whan Lee is a full professor at Korea University, where he is the head of the Department of Artificial Intelligence.

He received the B.S. degree in computer science and statistics from Seoul National University, Seoul, Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology in 1986 and 1989, respectively.

From February 1989 to February 1995, he was an Assistant Professor in the Department of Computer Science at Chungbuk National University, Cheongju, Korea. In March 1995, he joined the faculty of the Department of Computer Science and Engineering at Korea University, Seoul, Korea, and now he is the tenured full professor. In 2001, he stayed at the Department of Brain and Cognitive Sciences, MIT as a visiting professor.

A Fellow of the IAPR(1998), Korean Academy of Science and Technology(2009), and IEEE(2010), he has served several professional societies as chairman or governing board member. He was the founding Co-Editor-in-Chief of the International Journal of Document Analysis and Recognition and has been an Associate Editor of several international journals: Pattern Recognition, ACM Trans. on Applied Perception, IEEE Trans. on Affective Computing, Image and Vision Computing, International Journal of Pattern Recognition and Artificial Intelligence, and International Journal of Image and Graphics.

His research interests include pattern recognition, artificial intelligence, and brain engineering. He has more than 400 publications in international journals and conference proceedings, and authored 10 books.

As Deep Neural Networks (DNNs) have demonstrated superhuman performance in a variety of fields, there is an increasing interest in understanding the complex internal mechanisms of DNNs. However,despite the impressive performance, the adoption of DNNs is sometimes hindered by a transparency issue that arises from the complex internal structure of DNNs.

In this talk, I will introduce an explaining method: Relative Attributing Propagation (RAP), which decomposes the output predictions of DNNs with a new perspective of separating the relevant (positive) and irrelevant (negative) attributions according to the relative influence between the layers.

I will briefly describe the trend of deep learning, the necessity of explainability and well-known explaining methods. I explain the difference of purpose between these methods and present recent results of RAP.

To verify that the attributions propagated by RAP correctly account for the assumption, we utilize the evaluation metrics: (i) Outside-inside relevance ratio, (ii) Segmentation mIOU and (iii) Region perturbation.

I will also describe some success of explanation in real-world domain such as medical and video and the usage of applications.

Intelligent Systems and a Language Transparent World

Waibel, Alexander
Professor
Karlsruhe Institute of Technology & Carnegie Mellon University

Adenauerring 2, 76131 Karlsruhe, Germany & 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA
Alexander.waibel@kit.edu
Isl.anthropomatik.kit.edu & interact.kit.edu

Alexander Waibel is Professor of Computer Science at Carnegie Mellon University (USA) and at the Karlsruhe Institute of Technology (Germany). He is director of the International Center for Advanced Communication Technologies (interACT). The center now includes eight leading Universities around the world and coordinates joint research, faculty and student exchange programs between them.

Waibel is known for his work on AI, Machine Learning (ML), Multimodal Interfaces and Speech Translation Systems. He proposed early Neural Network based Speech and Language systems, including the TDNN, the first shift-invariant ("Convolutional") Neural Network, and early Neural Language systems. Based on advances in ML he and his team developed innovative multimodal interfaces (e.g., emotion recognition, face tracker, lipreading, error repair, meeting browser, smart rooms, human-robot collaboration). He also pioneered consecutive and simultaneous speech translation and multilingual communication systems that now break down language barriers.

Waibel founded more than 10 companies to transfer these academic results to practical deployment. This included "Jibbigo", the first speech translator on a phone (acquired by Facebook 2013), "Lecture Translator", the first automatic simultaneous translation for lectures (2005/2012), and KITES simultaneous transcription/ translation services (acquired by Zoom in 2021).

Waibel is a member of the German National Academy of Sciences, Leopoldina. He is a Fellow of IEEE and ISCA and recipient of the IEEE Flanagan Field Medal 2023. He published extensively (900+ papers, h-index 94) and received numerous awards. He received BS, MS and PhD degrees from MIT and CMU, respectively.

After centuries of separation and misunderstandings, we are lucky to be living in the generation that can see an end to language/cultural barriers between the peoples of our planet. AI, specifically: automatic recognition, translation and understanding of language in all its forms (text, speech, images, tactile, body, emotion etc.) is now making it feasible to operate and communicate in any country, language, mode and situation. We began working on first speech translation systems in the late 1980's on humble computing environments delivering modest performance, but have by now arrived at systems that are able to deliver real-time simultaneous interpretation capabilities for reallife situations that approach/exceed human performance. Key to success is a combination of 1.) better machine learning algorithms, 2.) exponential growth in available computing and data, and 3.) matching language systems with appropriate multimodal human-machine interfaces that make them usable.

In my talk, I will present different classes of systems that we have designed and deployed in different environments and for different requirements, including:

- Interpreting Smartphones for tourists, healthcare and relief workers
- Communicators on tablets in Humanitarian and Government Missions
- Road sign interpretation to translate international signing
- Multilingual subtitling and translation of TV broadcasts
- Automatic Interpretation of lectures for foreign students
- Symbiotic human & machine interpreters at the European Parliament
- Multimodal interpretation of human meta-communication (emotion, gaze…)
- Translation of silent speech
- Video translation and face dubbing

I will discuss the technology's progression and the lessons learned for each of these systems and show how language technology must meet user interface concerns. I will conclude with remarks on open challenges that need to be addressed to make language barriers truly transparent.

Explainable Artificial Intelligence: Can We Now Understand Internal Mechanisms of Deep Neural Networks?

Jaesik Choi
Associate Professor
Korea Advanced Institute of Science and Technology

Jaesik.choi@kaist.ac.kr

Prof. Jaesik Choi is an associate professor in the Graduate School of Artificial Intelligent at KAIST. He is a director of Explainable Artificial Intelligence Center established by Ministry of Science and ICT of Korea.

His research is concerned with statistical inference and machine learning for large-scale artificial intelligence problems including scaling up inference algorithms for large-scale dynamic systems, predictive analysis for time series data and its application to large-scale manufacturing systems. Some of his recent research results include the Lifted Relational Kalman Filtering (a scalable, linear time Kalman filtering algorithm), the Spatio-Temporal Pyramid Matching Kernel (the first pyramid matching kernel for spatiotemporal data), and the Relational Automatic Statistician (an automated, explainable interface for multivariate time series data).

Previously, he was an associate professor in the Electrical and Computer Engineering at UNIST until August 2019. Previously, he was an assistant professor at UNIST since July 2013. He was a Computer Scientist Postdoctoral Fellow of Computational Research Division at the Berkeley Lab. He received his Ph.D in Computer Science from University of Illinois at Urbana-Champaign in 2012 and received B.S. degree in Computer Engineering from Seoul National University in 2004.

He built an AI based automated control system for blast furnace of POSCO. The technology is selected as a national core technology by Ministry of Trade, Industry and as lighthouse factory by World Economy Forum. Prof. Choi received a prime minister's commendation for the research of relational automatic statistician and industrial application of time series deep learning models.

Currently, he is a POSCO fellow professor. He is a lifetime member of Korean Institute of Information Scientists and Engineers (KIISE) and a member of KIISE Artificial Intelligence Society.

Recent advances in artificial intelligence have changed our daily lives dramatically. Recently, a management consulting firm, McKinsey, expected that automation of knowledge work could impact $5-7 trillion worth of labor market across a wide range of industry sectors in 2025. In this talk, I will present several state-of-the-art machine learning framework to explain complicated artificial intelligent system.

First, I will introduce a method to explain multiple time series data. Analysis multiple time series data and explain the analyzed results in a human-readable form. Gaussian Processes (GPs) provide a general and analytically tractable way of modeling complex time-varying, nonparametric functions. The Automatic Bayesian Covariance Discovery (ABCD) system constructs natural-language description of time-series data by treating unknown time-series data nonparametrically using GP with a composite covariance kernel function. Unfortunately, learning a composite covariance kernel with a single time-series data set often results in less informative kernel that may not give qualitative, distinctive descriptions of data. We address this challenge by proposing two relational kernel learning methods which can model multiple time-series data sets by finding common, shared causes of changes. We show that the relational kernel learning methods find more accurate models for regression problems on several real-world data sets; US stock data, US house price index data and currency exchange rate data.

Second, I will present a method to analyze the inside of deep neural networks. Deep generative neural networks (DGNNs) have achieved realistic and high-quality data generation. In particular, the adversarial training scheme has been applied to many DGNNs and has exhibited powerful performance. Despite of recent advances in generative networks, identifying the image generation mechanism still remains challenging. In this talk, I will present an explorative sampling algorithm to analyze generation mechanism of DGNNs.

Socially-Interactive Artificial Intelligence: Perception, Synthesis and Learning of Human-Like Behaviors

Elisabeth André
Full Professor, Chair of Human-Centered Artificial Intelligence
Augsburg University

Universitätsstr. 6a, 86159 Augsburg, Germany
elisabeth.andre@uni-a.de
http://elisabethandre.de/

Elisabeth André is a full professor of Computer Science and Founding Chair of Human-Centered Artificial Intelligence at Augsburg University in Germany where she has been since 2001. Previously, she was a principal researcher at the German Research Center for Artificial Intelligence (DFKI GmbH) in Saarbrücken.

Elisabeth André has a long track record in multimodal human-machine interaction, social robotics, affective computing and social signal processing. Her work has won many awards including the ICMI Sustained Accomplishment Award 2021 and the Gottfried Wilhelm Leibnitz Prize 2021 of the German Research Foundation, with 2.5 Mio € the highest endowed German research award. Elisabeth André has served as a General and Program Co-Chair of major international conferences including ACM International Conference on Intelligent User Interfaces (IUI), ACM International Conference on Multimodal Interfaces (ICMI) or International Conference on Autonomous Agents and Multiagent Systems (AAMAS). In 2010, Elisabeth André was elected a member of the prestigious Academy of Europe, and the German Academy of Sciences Leopoldina. To honor her achievements in bringing Artificial Intelligence techniques to Human-Computer Interaction, 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. Since 2019, she is serving as Editor-in-Chief of IEEE Transactions on Affective Computing. Most recently, she was named one of the ten most influential figures in the history of AI in Germany by National Society for Informatics (GI).

The automatic analysis and synthesis of social signals conveyed by voice, gestures, mimics, etc., will play a vital role for next-generation interfaces as it paves the way towards a more intuitive and natural human-computer interaction with robots and virtual agents. In my talk, I will present computational methods to implement socially interactive behaviors in artificial agents, focusing on three essential properties of socially interactive interfaces: Social Perception, Socially-Aware Behavior Synthesis, and Learning Socially-Aware Behaviors. Besides analytic methods informed by theories from the cognitive and social sciences, I will discuss empirical approaches that enable an artificial agent to learn socially interactive behaviors from recordings of human-human interactions or life interactions with human interlocutors. I will highlight opportunities and challenges that arise from neural behavior generation approaches that promise to achieve the next level of human-likeness in virtual agents and social robots. Finally, I will share lessons we learnt during the development of socially interactive agents. To benefit users, we do not just have to work on technical solutions, but go beyond disciplinary boundaries to encompass ethical, legal, and social implications of employing such agents.

Conversation Modeling for Computational Social Science

Alice Oh
Associate Professor
Korea Advanced Institute of Science and Technology

alice.oh@kaist.edu
https://aliceoh9.github.io/

Alice Oh is an Associate Professor in the School of Computing at KAIST. She received her PhD in 2008 from MIT and joined KAIST in the same year. Her major research area is at the intersection of machine learning and computational social science.

Within machine learning, she studies various models designed for analyzing written text including social media posts, news articles, and personal conversations. She also looks at non-textual data such as social network friendship and logs from online games for which she interacts closely with social scientists for an interdisciplinary approach to computational social science. A particular application focus of applying computational methods to a social science problem is computer science education. Her students have developed a Web-based system for improving programming education, and through that system they collect and analyze large-scale, fine-grained student behavior data.With that data, they aim to understand the behaviors of students and teaching assistants via machine learning models such that they can offer identification of students in need of assistance, provide automatic assistance for simple problems, track students' progress, and help students to learn better through social learning.

In this talk, I will present research results from three recent papers on conversational modeling. The first paper looks at a problem in Korean history and applies a machine learning-based approach. We make a corpus of historical records of the Joseon Dynasty to predict the rulers' decision-making styles using a model based on the hierarchical attention network with an addition of speaker embedding. The second paper tackles a problem in clinical psychology. We build an annotated corpus of text-based counseling conversations and classify the client utterances into clinically meaningful categories. We then develop a model based on ULMFit, but with modifications for classifying utterances in a conversation. Last, we look at the problem of dialogue response generation. We build a corpus of Twitter conversations in which many users carry on multiple longterm conversations with multiple partners. We then develop a variational hierarchical model with speaker embedding which shows significantly improved performance for response generation, especially in cases of unseen users. These three papers make meaningful progress in the field of computational social science using conversation modeling.

Memo

Memo

The 6th KAST-Leopoldina Bilateral Symposium

Artificial Intelligence and the Digital Age - Implications for the Future of Society

SESSION 2

THE CRUCIAL ROLE OF DATA

Co-chair

Seong-Whan Lee (Korea University)

Alexander Waibel (Karlsruhe Institute of Technology and Carnegie Mellon University)

Speakers

Adversarial Learning to Mitigate Algorithmic Bias of AI Systems Caused by Data

Sung-Bae Cho (Yonsei University)

Self-Learning Based Defect Reduction for Dialogue Systems

Eunah Cho (Amazon)

On the Role of Public Biomedical Data for Personalized Medicine

Hyunjung Shin (Ajou University)

Adversarial Learning to Mitigate Algorithmic Bias of AI Systems Caused by Data

Sung-Bae Cho
Professor
Yonsei University

sbcho@yonsei.ac.kr

Sung-Bae Cho received the Ph.D. degree in computer science from KAIST (Korea Advanced Institute of Science and Technology), Taejeon, Korea, in 1993. He was an Invited Researcher of Human Information Processing Research Laboratories at Advanced Telecommunications Research (ATR) Institute, Kyoto, Japan from 1993 to 1995, and a Visiting Scholar at University of New South Wales, Canberra, Australia in 1998. He was also a Visiting Professor at University of British Columbia, Vancouver, Canada from 2005 to 2006, and at King Mongkut's University of Technology Thonburi, Bangkok, Thailand in 2013. Since 1995, he has been a Professor in the Department of Computer Science, Yonsei University, Seoul, Korea.

His research interests include hybrid intelligent systems, soft computing, evolutionary computation, neural networks, pattern recognition, intelligent man-machine interfaces, and games.

Dr. Cho has been serving as an associate editor for several journals including IEEE Transactions on CI and AI on Games (2009-present) and IEEE Transactions on Fuzzy Systems (2013-present). He was also the chair of Games Technical Committee, IEEE CIS (2009-2010), and Student Games-based Competition Subcommittee, IEEE CIS (2011-2012). He is a member of Board of Government (BoG) of Asia Pacific Neural Networks Society (APNNS) (2011-present), and a member of three technical committees in IEEE CIS such as Emergent Technologies, Computational Finance and Economics, and Games.

Dr. Cho has been awarded several best paper prizes from IEEE Korea Section (1990), Korea Information Science Society (1993, 2005), International Conference on Soft Computing (1996, 1998), World Automation Congress (1998), International Conference on Information Networking (2001), and International Conference on Hybrid AI Systems (2011). He was also the recipient of the Richard E. Merwin prize from IEEE Computer Society in 1993.

Currently, he is the Vice President of the Korea Cognitive Science Society, the Korea Brain Engineering Society, and the Korea Information Science Society.

Discrimination is the unfair treatment of individuals based on sensitive attributes such as gender and race. It has been found that machine learning, which has significantly led to constructing a model capable of deciding the labels of novel data, can lead to unexpected results with bias. Even machine learning algorithms have amplified algorithmic bias, but it cannot be solved by removing the sensitive variables from the data. For example, even if we delete race information, it is possible to derive race by zip code. There are several methods to mitigate the bias: pre-processing, inprocessing, and post-processing. Pre-processing is to solve the problem by eliminating the bias present in the training data itself, in-processing is to reduce the bias by adding a constraint to the learning algorithm even if there is a bias in the data, and post-processing is to ensue decisions themselves.

In this talk, non-discriminated representation is formulated as a dual objective optimization problem of encoding data while obfuscating the information about the protected features in the data representation by exploiting the unbiased information bottleneck. Encoder learns data representation and discriminator judges whether there is information about the protected features in the data representation or not. They are trained simultaneously in adversarial way to achieve fair representation. Moreover, the algorithmic bias is analyzed in terms of bias-variance dilemma to reveal the cause of bias, so as to prove that the proposed method is effective for reducing the algorithmic bias in theory and experiments.

Experiments with the well-known benchmark datasets such as Adults, Census, and COMPAS demonstrate the efficacy of the proposed method compared to other conventional techniques. Our method not only reduces the bias but also can use the latent representation in other classifiers (i.e.,once a fair representation is trained, it can be used in various classifiers). We illustrate it by applying to the conventional machine learning models and visualizing the data representation with t-SNE algorithm.

Self-Learning Based Defect Reduction for Dialogue Systems

Eunah Cho
Senior Applied Scientist
Amazon, Alexa AI

10450 NE 10TH ST BELLEVUE, WA, 98004
eunahch@amazon.com

Dr Eunah Cho is currently a Senior Applied Scientist in Alexa natural language understanding, focusing on automated learning from user feedback for dialogue systems. Before joining Amazon, Dr Cho was a Post-Doctoral Researcher at the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, conducting research in the field of machine translation for spoken language. Prior to her PhD at KIT, she obtained an MSc. degree at Aalto University. Her research interests include deep learning for natural language processing and life-long learning efforts such as semi-supervised learning, few-shot learning, and data augmentation/generation. She is currently an organizing committee member of the life-long learning workshop.

Query rewriting (QR) is an increasingly important technique for reducing user defect in a conversational AI system. User defect is caused by various reasons, including errors in automatic speech recognition (ASR), natural language understanding (NLU), entity resolution (ER) component, or users' slip of the tongue. Often the user defects stem from personalized factors as well, such as user's speech pattern, dialect, or preferences. In this talk, I will present a search-based self-learning QR framework, which focuses on automatic reduction of user defect for large scale conversational AI agents. I will present the QR framework applicable globally as well as the one focusing on personalized defect reduction. Supported by user feedback-based learning, training our models does not require hand-annotated data. I will demonstrate the effectiveness our QR system for both global and personalized level, through offline and online experiment on user traffic.

On The Role of Public Biomedical Data for Personalized Medicine

Hyunjung (Helen) Shin
Professor
Dept. of Industrial Engineering / Artificial Intelligence / Integrative System Engineering
Ajou University, South Korea

shin@ajou.ac.kr

Hyunjung (Helen) Shin is a professor at Ajou University in South Korea. She received the Ph.D. degree in Data Mining from Seoul National University (SNU), Seoul, South Korea in 2005, and further worked on Machine Learning and Bioinformatics as a research scientist at Max Planck Institute (MPI), Tuebingen in Germany (2004-2006).

After returning from Germany, she worked at Seoul National University College of Medicine as a research professor. Since 2006, she joined Ajou University in Korea as a faculty member of the Department of Industrial Engineering.

Currently, she provides academic services as a member of board of directors in Business Intelligence & Data Mining Society (BIDM), Korean Institute of Information Scientists and Engineers (KIISE), Artificial Intelligence Society at KIISE, Korean Institute of Industrial Engineers (KIIE), Translational Bioinformatics Conference (TBC), and Korean Society for Bioinformatics and Systems Biology (KSBSB). Also, she has been the vice chairman for Billing Software Inspection and Review Committee of Health Insurance Review and Assessment Service (HIRA).

Her theory interest is focused on Machine Learning algorithms, particularly in Network Learning methods. Her research activities range across diverse areas including network analytics, protein function and cancer genotype-phenotype prediction, multi-omics data integration, hospital fraud detection, etc.

In this talk, we present a sketch of precision medicine applications using personal body-oriented secrecies plugging into the backbone built by public biomedical big data. As a case, we introduce Gene Ranker that produces scores for target genes. It accommodates both numerous general genes and disease-specific genes. Gene Ranker employs the protein-protein interaction as a backbone network. And, disease-specific information is added onto it in a form of a weighted gene coexpression network of patients with the specified immune disease. Then the general gene network takes on peculiarities of the specified disease. Another case is on dementia. By incorporating personal SNPs, one can obtain his own gene scores for Alzheimer, Vascular dementia, and Parkinson's diseases. The implementations of those work were all geared with the network-based machine learning algorithms and were fueled by various types of genomic and diseasomic big data. The validation was conducted by text-mining existing evidences from approximately 28 million PubMed literature.

Memo

Memo

The 6th KAST-Leopoldina Bilateral Symposium

Artificial Intelligence and the Digital Age - Implications for the Future of Society

SESSION 3

FROM APPLICATION TO ACCEPTANCE

Co-chair

Chang Dong Yoo (KAIST)
Alexander Waibel (Karlsruhe Institute of Technology and Carnegie Mellon University)

Speakers

AI Methods for Predicting Toxicity of Compounds in Chemical and Genetic Space
Sun Kim (Seoul National University)

Autonomy and Trust in Self-improving AI Systems
Byoung-Tak Zhang (Seoul National University)

How to Create Successful Technology Disruption
Alexander Waibel (Karlsruhe Institute of Technology and Carnegie Mellon University)

Fair AI
Chang Dong Yoo (KAIST)

Constructivist Psychology and the Digital Age
Reinhold Kliegl (University of Potsdam)

AI for Social Impact: Poverty and Disaster Mapping from the Sky
Meeyoung Cha (KAIST)

Interfaces of AI and Democracy
Jeanette Hofmann (Freie Universität Berlin)

AI Methods for Predicting Toxicity of Compounds in Chemical and Genetic Space

Sun Kim
Professor, Computer Science and Engineering
Director, Bioinformatics Institute
Seoul National University

Sunkim.bioinfo@snu.ac.kr

Sun Kim is Professor in the School of Computer Science and Engineering and an adjunct faculty of Biological Sciences at Seoul National University. He is also President of Mogam Institute for Biomedical Research which is being completely transformed to AI-based Drug Discovery Institute. Before joining SNU, he was Chair of Faculty Division C; Director of Center for Bioinformatics Research, an Associate Professor in School of Informatics and Computing at Indiana University (IU) Bloomington. Prior to joining IU in 2001, he worked at DuPont Central Research from 1998 to 2001, and at the University of Illinois at Urbana-Champaign from 1997 to 1998. Sun Kim received B.S and M.S and Ph.D in Computer Science from Seoul National University, KAIST and the University of Iowa, respectively.

Sun Kim is a recipient of Outstanding Junior Faculty Award at Indiana University 2004, US NSF CAREER Award DBI-0237901 in 2003. He is actively contributing to the bioinformatics community, serving on the editorial board for journals including editors for the METHODS journal and International Journal of Data Mining and Bioinformatics, associate editor-in-chief for ACM/IEEE Transactions on Computational Biology and Bioinformatics, having served on the board of directors for ACM SIG Bioinformatics and for education for the IEEE Computer Society Technical Committee on Bioinformatics. He organized many scientific meetings including IEEE BIBM 2008, 2009, 2020, APBC 2020, ACM BCB 2011. Among many service activates in Korea, he served on Samsung Future Technology Committee for 2016-2018, a member of The National Science and Technology Council (NSTC) of the Korean Government for 2019-2020, President of Korea Artificial Intelligence Society (2016-2018) and Vice President of Korea Society of Bioinformatics and Systems Biology (2011 - Present).

Toxicity is a major issue in developing drug and environmental materials. The scientific community accumulated quite amount information in terms of compound structure information and also in terms of genetic response of toxic compounds in cells. Thus, we have now an o2pportunity to utilize the data to model toxicity of compounds using AI technologies in chemical and genetic space. Predicting toxicity in chemical space requires mining substructures that can be characterized as toxicity, a.k.a., toxicophore. The challenge in identifying toxicophores is to enumerate and mine substructures for toxicity from numerous substructures of known toxic compounds. We used two different approaches, one based on subgraph mining and another leveraging positive unlabeled (PU) learning on graphs. Defining toxic signature in the genetic space is much more challenging since cellular response to toxic compounds depends on the dosage and exposure time to toxic compounds. Gene expression data or transcriptomic profile upon drug treatment has been accumulated in several databases including Library of Integrated Network-based Cellular Signatures (LINCS) for more than a decade. However, it is very difficult to identify which of transcriptome profiles represent response to toxic drugs since cellular response depends on the dosage and exposure time after toxic drug treatment. We used one-class boundary prediction technique to determine toxic transcriptomic profiles and stratified network-based distance or "network ruler" to determine most likely toxic profiles. In a cross validation with two non-overlapping data sets, we were successful in matching toxic profiles defined by two independent network rulers.

Autonomy and Trust in Self-improving AI Systems

Byoung-Tak Zhang
Professor of Computer Science and Engineering,
Seoul National University

btzhang@bi.snu.ac.kr

Byoung-Tak Zhang is POSCO Chair Professor of Computer Science, Cognitive Science, and Brain Science at Seoul National University (SNU) and directs AI Institute (AIIS), the Cognitive Robotics and Artificial Intelligence Center (CRAIC). He currently serves as President of the Korean Society for Cognitive Science (KSCS) and Chairman of the National Association of Cognitive Science Industries (NACSI) and had served as President of the KIISE Society for Artificial Intelligence (2010-2013). He received his Ph.D. in computer science from University of Bonn, Germany, in 1992, and his B.S. and M.S. in computer science and engineering from Seoul National University in 1986 and 1988, respectively.

His research questions center around: i) How does the brain build a model of the world to act so fast, flexibly, and robustly, and ii) How we can build an intelligent machine that can see, talk, reason, act, and learn like a brain in a real world. He has been investigating brain-inspired cognitive learning architectures and algorithms for artificial intelligence and cognitive science. He has developed the deep hypernetwork models of cognitive learning and memory and applied them to language, music, video, and robot learning. Recently, his AUPAIR robot team won the 2017 RoboCup@Home (Social Standard Platform League) in which humanoid robots should complete missions in real-world-like environments such as home, cocktail bar, and restaurants.

Before joining Seoul National University in 1997, he has worked as Research Fellow at the German National Research Center for Information Technology (GMD, now Fraunhofer Institutes) in Sankt Augustin/Bonn for 1992-1995.

He has been Visiting Professor at MIT CSAIL and Brain and Cognitive Sciences Department, Cambridge, MA for 2003-2004, Samsung Advanced Institute of Technology (SAIT) for 2007-2008, BMBF Excellence Centers of Cognitive Technical Systems (CoTeSys, Munich) and Cognitive Interaction Technology (CITEC, Bielefeld) for the Winter of 2010-2011, and Princeton Neuroscience Institute (PNI) for 2013-2014.

He serves as Associate Editor of Journal of Cognitive Science, Applied Intelligence, BioSystems, and the IEEE Transactions on Evolutionary Computation (1997-2010).

He has received numerous awards and honors, including Red Stripes Order of Service Merit, INAK Award, Minister of Science and Technology Award, Okawa Research Grant Award, Distinguished Service Award from the IEEE Computational Intelligence Society, and Academic Excellence Award from the Korea Information Science Society

Machine learning has revolutionized AI by making machines self-improve their performance based on observed data. Recent learning AI systems begin to improve themselves more actively by choosing or synthesizing their own training data. When equipped with sensors and actuators, the machines can even fully autonomously learn by generating new data by self-experimentation. What if the self-improving AI runs against the purpose or goes rogue to do harm? How can we make the autonomously self-improving AI systems still trustworthy, safe, and secure?

How to Create Successful Technology Disruption

Waibel, Alexander
Professor
Karlsruhe Institute of Technology & Carnegie Mellon University

Adenauerring 2, 76131 Karlsruhe, Germany & 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA
Alexander.waibel@kit.edu
Isl.anthropomatik.kit.edu & interact.kit.edu

Alexander Waibel is Professor of Computer Science at Carnegie Mellon University (USA) and at the Karlsruhe Institute of Technology (Germany). He is director of the International Center for Advanced Communication Technologies (interACT). The center now includes eight leading Universities around the world and coordinates joint research, faculty and student exchange programs between them.

Waibel is known for his work on AI, Machine Learning (ML), Multimodal Interfaces and Speech Translation Systems. He proposed early Neural Network based Speech and Language systems, including the TDNN, the first shift-invariant ("Convolutional") Neural Network, and early Neural Language systems. Based on advances in ML he and his team developed innovative multimodal interfaces (e.g., emotion recognition, face tracker, lipreading, error repair, meeting browser, smart rooms, human-robot collaboration). He also pioneered consecutive and simultaneous speech translation and multilingual communication systems that now break down language barriers.

Waibel founded more than 10 companies to transfer these academic results to practical deployment. This included "Jibbigo", the first speech translator on a phone (acquired by Facebook 2013), "Lecture Translator", the first automatic simultaneous translation for lectures (2005/2012), and KITES simultaneous transcription/ translation services (acquired by Zoom in 2021).

Waibel is a member of the German National Academy of Sciences, Leopoldina. He is a Fellow of IEEE and ISCA and recipient of the IEEE Flanagan Field Medal 2023. He published extensively (900+ papers, h-index 94) and received numerous awards. He received BS, MS and PhD degrees from MIT and CMU, respectively.

The successes and failures of cultures and societies have historically been accompanied if not triggered by technology disruption that moved the playing field and shifted the balance of power, rather than by incremental improvements in playing the game. But how can such disruption be predicted or even organized and managed to the benefit of society?

That it is better to disrupt internally than to catch up with external innovations, was the simple insight understood by John F. Kennedy, when he set up the United States "Defense Advanced Research Project Agency" (DARPA) in response to Sputnik, a technology surprise presented by the USSR, a formidable opponent at the time. DARPA was exceptionally successful among research project agencies and went on to pioneer many technologies that define our world today: The Personal Computer, the Internet, GPS, Speech Reco, Machine Translation, and many more. It triggered new businesses and industries. Envied, but rarely successfully copied, DARPA had many wondering: How was it able to manage such a feat? The answers are complex, but follow clear principles.

In this talk, I will discuss the organizational principles of DARPA that enabled this success, and I will discuss whether a similarly successful agency for disruptive research is possible elsewhere and around different goals and needs.

Fair AI

Chang Dong Yoo
Professor & Adjunct Professor
School of Electrical Engineering & School of Computer Science
KAIST

cd_yoo@kaist.ac.kr
http://slsp.kaist.ac.kr

Chang D. Yoo is a professor in the School of Electrical Engineering at Korea Advanced Institute of Technology (KAIST) and Adjunct Professor in the Department of Computer Science. He has been serving as the president of the Korean AI Association since April 2019. He is currently the Director of two government-sponsored AI centers with over 100 participating members. He received the B.S. degree in Engineering and Applied Science from the California Institute of Technology , the M.S. degree in Electrical Engineering from Cornell University and the Ph.D degree in Electrical Engineering from the Massachusetts Institute of Technology. From January 1997 to March 1999 he was Senior Researcher at Korea Telecom (KT). He also served as Dean of the Office of Special Projects and Dean of the Office of International Relations, respectively. Dr. Yoo has been a consultant for many firms and was a consultant for the Korean Foundation for Advance Studies (one of the oldest and largest scholarship foundations in Korea). He is also Director of Korea Institute of Electrical Engineers (KIEE) and Director of the Acoustical Society of Korea (ASK). From March 2005 to March 2006, he was a visiting scholar to the Research Laboratory of Electronics (RLE) at MIT. In March 2015, he returned to MIT and spent another sabbatical year at RLE. He is Member of Tau Beta Pi and Sigma Xi. He was on the technical committee member of IEEE machine learning for signal processing society from 2009 to 2011. He had also served as Associated Editor of IEEE Signal Processing Letters (2011-2012), IEEE Transactions on Information Forensics and Security (2012-2013), IEEE Transactions on Audio, Speech and Language Processing (2011-2014). From Nov. 2015, he has been on the Editorial Board of the Journal of Signal Processing Systems. He has taught Digital Signal Processing, Digital Speech Processing, Machine learning for a number of years. His current research interests include the application of digital signal processing and machine learning theories in multimedia processing and digital communications. Dr. Yoo was born in Bonn Germany and grew up in Pakistan, Korea, US and UK.

It is estimated that within 10 years, Artificial Intelligence (AI) will have found its way into nearly all sectors of our lives, both personal and professional. While the technology has many benefits— ranging from enhanced life style to public service—it is also a source for concern. The growing use of artificial intelligence in sensitive areas, including for hiring, criminal justice, and healthcare, has stirred a debate about bias and fairness. This talk will look at the issue of fairness in AI, and introduce a number of efforts made around the globe including at the AI Fairness Research Center at KAIST.

Constructivist Psychology and the Digital Age


© Karla Fritze

Reinhold Kliegl
Senior Professor of Psychology
University of Potsdam

Am Neuen Palais 10, 14469 Potsdam, Germany
kliegl@uni-potsdam.de
https://www.uni-potsdam.de/en/trainingswissenschaft/staff/rkliegl.html

Reinhold Kliegl is senior professor of psychology at the University of Potsdam, Germany. His past research focused on how the dynamics of language-related, perceptual, and oculomotor processes subserve attentional control, using reading, spatial attention, and working memory tasks as experimental venues; he also examined neural correlates and age-related differences in these processes. His research has been carried out in interdisciplinary projects with colleagues from linguistics as well as from theoretical physics, mathematics, and sports science/medicine. His current research as senior professor focuses on the simultaneous modeling of experimental effects and individual differences related to socially relevant issues. Two examples are the physical fitness of children and the cognitive processing of social-media information. He is a member of Berlin-Brandenburg Academy of Sciences and of Deutsche Akademie der Naturforscher Leopoldina (German National Academy of Sciences). Major honors: Gottfried Wilhelm Leibniz Prize of Deutsche Forschungsgemeinschaft (2002), Wilhelm-Wundt-Medaille and Honorary Membership of Deutsche Gesellschaft für Psychologie (2008); Fellow of the Association for Psychological Science (2011); Lifetime Achievement Award of Deutsche Gesellschaft für Psychologie (2020).

Our phenomenal world (i.e., the world as we experience it) is to a large degree a construction of the mind, moderated to varying degrees by sensory information. Thus, behavior is driven by knowledge- and experience-based expectations and predictions, not unlike AI algorithms. Both the degree to which our phenomenal world agrees with the objective world and the degree to which we are aware of these discrepancies vary widely. These discrepancies are foundational for an understanding of and susceptibility to manipulation in the context of social media. For example, some people want to protect themselves against "fake news" or even engage to counter them, but others use them for better or worse (e.g., cherish social divisiveness). And it is difficult to discern to what degree discrepancies between objective (intersubjective) and subjective reality are consciously accessible (e.g., used with a purpose in an instrumental fashion to achieve a political goal) or inaccessible to conscious reflection (e.g., closer to hallucinations or cult experience). Adopting signal detection theory for the separation of discriminability and response bias will be introduced to address a core psychological issue in this context: confirmation bias. Some core AI issues related to this human condition are the generation of synthetic media (e.g., deepfakes), microtargeting (e.g., in political campaigns), and the explainability of algorithmic decisions. These issues are socially transformative because the use of usually unrepresentative information harvested from the web and especially from social media appears to exacerbate the discrepancies between individuals' (or even entire social groups') phenomenal worlds. To strike a constructive balance between transparency and privacy, we need regulations (e.g., audits, IRBs) to prevent mal-use and assure safe and responsible usage of AI-based algorithms.

AI for Social Impact: Poverty and Disaster Mapping from the Sky

Meeyoung Cha
Associate Professor & Chief Investigator
Korea Advanced Institute for Science and Technology & Institute for Basic Science

mcha@ibs.re.kr

Meeyoung Cha is an associate professor in the School of Computing at Korea Advanced Institute of Science and Technology (KAIST) and an adjunct professor in the Department of Brain and Cognitive Sciences and Graduate School of Culture Technology. She previously worked as a post-doctoral researcher at the Max Planck Institute for Software Systems (MPI-SWS) in Saarbrücken, Germany.

Meeyoung's interests include data science and information science, with a focus on modeling socially relevant information propagation processes. Her research on misinformation, poverty mapping, fraud detection, and long-tail content has received over 18,000 citations and best paper awards at a number of conferences. She has received the Korean Young Information Scientist Award 2019, the AAAI ICWSM Test of Time Award 2020, and the Minister's Award of Science and ICT of Korea 2022.

Meeyoung has worked as a visiting professor at Facebook's Data Science Team in Menlo Park, California, and has been named the World Customs Organization (WCO)'s BACUDA science collaborator. She is a member of the Seoul Forum for International Affairs (SFIA) and a commissioner for the Korea Copyright Commission, the Korea Customs Service, the National Tax Service, the Open Data Mediation Committee (ODMC), and the Presidential Council on Intellectual Property. As a Chief Investigator, she also leads the Data Science Research Group at the Institute for Basic Science (IBS) in Korea.

Artificial intelligence (AI) is reshaping business and science. One of the areas it has an impact on is achieving the Sustainable Development Goals (SDGs). This talk will review some of the latest research advances on poverty mapping (goal #1) and climate action (goal #13). I will discuss the problem of inferring economic development in the developing world with few official statistics. One emerging technology is to use deep image learning on high-resolution daytime satellite imagery. The same technology can be used to detect disaster damage at the building level, assisting in the rapid response. I'll conclude the talk by discussing other exciting opportunities for using data science and AI for social impact.

Interfaces of AI and Democracy


© David Ausserhofer

Jeanette Hofmann
Professor of Internet Policy
Freie Universität Berlin

Garystr. 55, 14195 Berlin
jeanette.hofmann@wzb.eu

Jeanette Hofmann has a background in political science and teaches at the Institute for Media and Communication Studies of the Freie Universität Berlin. She heads the research group "Politics of Digitalisation" at the Berlin Social Science Center (WZB); she is a founding co-director of the Alexander von Humboldt Institute for Internet and Society, and principal investigator at the Weizenbaum Institute for the Networked Society.

Professor Hofmann's current research addresses the interplay of democracy and digitalisation with a focus on algorithmic systems, the social shaping of digital technologies and the regulation of digital platforms. Her research groups study the politics of digitalization, the digital transformation of democracy and the logics of automated decision-making systems.

In addition to her academic duties, Prof Hofmann has served in various advisory capacities. She chaired an academic expert commission on digitalisation and democracy (German National Academy of the Sciences Leopoldina), an academic commission on 'Young Engagement in the Digital Age' (German Government), and she was a member of the expert group to the EU Observatory on the Online Platform Economy.

My presentation addresses the future relationship between machine learning technologies and democracy. More precisely, I want to discuss how machine learning and democracy relate to and may influence each other. This is not an easy task because the protagonists of machine learning development and democratic theory rarely meet and talk to each other. While 'politicking' has a bad reputation in computer science and most engineers prefer to keep their distance from the realm of politics, political scientists, in turn, are expecting predominantly bad things from AI. A major concern at present is that AI could undermine democratic ways of opinion formation and decision-making. What if algorithmic systems soon offer better decisions than democratically legitimised institutions?

It seems unhelpful to regard machine learning and democracy as categorically different, competing or even opposing systems. Such notions drive fears of technical domination and subjugation. Instead, in order to understand the interplay between AI and democracy we need to identify interfaces between the calculative logics of machine learning and the conflicting, majoritarian logics of democratic governance. One of these interfaces is the public sphere, which may encourage a mutual integration of norms and principles. For example, the current debate about the quality and impact of machine learning systems on society aims to affect the development of algorithmic systems through calls for greater accountability ("explainability") and for compliance with human rights principles such as equality and non-discrimination. However, reverse forms of influence are also likely. Machine learning-based norms and principles may transform our understanding of society and democracy. I will use the cybernetic principle of feedback loops as an example for illustrating this point in a somewhat speculative way.

Memo

Memo

Memo