중앙대학교 인문콘텐츠연구소
Humanities Research Institute, Chung-Ang University

ICAIH 2020

THE 3rd INTERNATIONAL CONFERENCE ON AI HUMANITIES

DATE Dec 8, 2020 17:00 (KST) OPENING PLENARY
Dec 8, 2020 – Dec 14, 2020

VENUE Chung-Ang University, South Korea
Online aihumanities.org/aih2020

SCHEDULE *ICAIH 2020 The 3rd International Conference on AI Humanities*

December 8, 2020 (Tue) 17:00 – 18:10 KST

OPENING PLENARY

PARTICIPATION ZOOM online meetings *Guidance and links to be sent later*
Korean-English/English-Korean simultaneous interpretation provided

	PROGRAMS	
	Moderator	Heeryon Cho Humanities Research Institute(Korea)
17:00 – 17:05 KST	Opening Speech	Chankyu Lee Humanities Research Institute(Korea)
17:05 – 17:15 KST	Introduction and Guidance on the Conclusion of MOU	Hyeongjoo Kim Humanities Research Institute(Korea)
17:15 – 17:25 KST	Speech of Congratulation and Introduction of International Journal JAHR	Amir Muzur Rijeka University(Croatia)
17:25 – 18:05 KST	Greetings and Introductions	All Speakers
18:05 – 18:10 KST	Closing Speech	

December 8, 2020 (Tue) 17:00 KST – December 14, 2020 (Mon) 24:00 KST

ICAIH 2020 CONFERENCE

PARTICIPATION ICAIH 2020 webpage (aihumanities.org/aih2020)
Humanities Research Institute Webpage (aihumanities.org)

INQUIRIES aihumanities2018@gmail.com

The 3rd International Conference on AI Humanities

KEYNOTE SPEECH

AMIR MUZUR RIJEKA UNIVERSITY, CROATIA AI and Empathy: Living in Perfect Harmony?

SESSION 1

HEAD OF THE SESSION Yong E. Rhee Korea University

HUMAN BEYOND MECHANICAL INTELLIGENCE	
Christian Byk Intergovernmental Bioethics Committee, UNESCO(France)	UNESCO'S work in the Field of Artificial Intelligence and the Concept of Transhumanism
Ivan Štajduhar Rijeka University(Croatia)	The Reach and Limitations of Machine Intelligence
Mariusz Pisarski University of Ss. Cyril and Methodius in Trnava(Slovakia)	Breakfast with Plato, Dinner with Queen Min: Linguistic Avatars of GPT-3
Olga Markic University of Ljubljana(Slovenia)	Second wave AI and new cognitive technologies: the question of transparency and trust
Robin Kopecký Karel Capek Center, Charles University(Czech Republic)	How Young Minds See Future Minds: Uncanny Valley and Ascription of Different Types of Attributes to Robots
Aditi Vashistha University of Delhi(India)	The Emotive Machine, and Mechanised Self
Jiseon Park Hankuk University of Foreign Studies(Korea)	Heidegger's Tools and Technology Criticism & the Era of Artificial Intelligence
Seonjin Yang Chung-ang University(Korea)	Human dignity and surveillance society in the age of artificial intelligence: centering on Jeong, je-du, Wang, Yang-Ming and H. Begson, M. Merleau-Ponty

SESSION 2

HEAD OF THE SESSION Cheongho Lee Sangmyung University

THE PROMISE AND PERILS IN THE ERA OF TECHNOLOGY	
Kuei Tien Chou Ngerng Yi-Ling National Taiwan University Risk Society and Policy Research Center (Taiwan)	Using the Kübler-Ross Model and a Proactive Approach Toward Managing Digital Fear by Adopting Iterative and Participatory Strategies on a Research-Literacy-Regulatory Spectrum
Tomáš Hříbek Czech Academy of Sciences (Czech Republic)	Human Work in the Digital Economy
Alfonso Ballesteros Universidad Miguel Hernández de Elche(Spain)	Digitocracy: Ruling and Being Ruled
Avantika Tewari Jawaharlal Nehru University(India)	Fighting the Machine or to Fight Capitalism?
Suhyoun Hwang Sogang University(Korea)	Can Artificial Intelligence Promote Consumption? Comparison of 'Say's Law and 'Demand Control'
Sanghyong Shin Andong National University(Korea)	The Implications of ARTIFICAL INTELLIGENCE
Yeonjae Kim Kongju Natinal University(Korea)	A Study on the Values and Humanism of Confucian Society from the Social Engineering Perspectives of Artificial Intelligence

SESSION 3

HEAD OF THE SESSION Insok Ko Inha University

REVOLUTIONIZE AI	
David Černý Karel Capek Center(Czech Republic)	Autonomous Vehicles, Discrimination, and Age
Filippo Santoni de Sio Delft University of Technology(Netherlands)	Moral control over AI, What it is and why it matters.
Jeffrey Chan Kok Hui Singapore University of Technology and Design(Singapore)	The socio-technical design of trustworthy AI
John Danaher National University of Ireland(Ireland)	AI and the Constitutions of the Future: How can AI revolutionize our social-moral order?
Luisa Damiano University of Messina(Italy)	Artificial Empathy in Socially Assistive Robotics. A Philosophical Exploration
Saša Horvat Rijeka University(Croatia)	Love all, trust a few–on human relationship with AI

SESSION 4

HEAD OF THE SESSION Eunjoo Choi Konkuk University

HUMAN AND NON-HUMAN CITY	
Iva Rinčić · Amir Muzur Rijeka University(Croatia)	Forget being (only) smart and start being (bio)ethical! Weighing pros and cons of technology use in urban settings
Srećko Gajović University of Zagreb(Croatia)	Zombification of humanity - a pandemic related phenomenon
Taesik Kim Masaryk University(Czech Republic)	Walking in the City in the Age of Artificial Intelligence
Aditya Nayak Jawaharlal Nehru University(India)	A.I. and Mental-Health: A Paradigm of Human Mechanisation

SESSION 5

HEAD OF THE SESSION Jae-uk Choo Chung-Ang University

THE WORLD PROCEEDED BY ARTIFICIAL INTELLIGENCE	
Dobrivoje Lale Eric Center for the Promotion of Science(Serbia)	Intelligence IO
Pak-Hang Wong University of Hamburg(Germany)	Artificial Intelligence, Personal Decisions, Consent, and the Confucian Idea of Oneness
Thomas Sukopp University of Siegen(Germany)	An argument against the unlimited applicability of Artificial Intelligence in classroom settings
Geonwoo Park Konkuk University(Korea)	Does Knowledge Gap Theory Explain the Difference?: An Empirical Study of Artificial Intelligence and Its Impact on Human's Perception
Heeryon Cho Humanities Research Institute(Korea)	AI and Revisiting the Two Cultures
Myeonghyeon lee Yujin Kang Chung-ang University(Korea)	Development and Application of <Tokki-jeon Chatbot> for Education Using Chatbot Open Source

SESSION 6

HEAD OF THE SESSION Ju-hee Eom Konkuk University

CALCULATING GOOD AND EVIL WITH ARTIFICIAL INTELLIGENCE	
Anita Sue Jwa Stanford University(U.S.A)	When Artificial Intelligence Meets Neuroscience: Social, Ethical, and Legal Implications of AI-powered Neurotechnology
Corinna Bath Braunschweig University(Germany)	Gender, bias and knowledge politics: towards new materialist reconfiguration of AI systems
Hortense Gallois McGill University, Center for Genomics and Policy(Canada)	Integrating AI into Healthcare: A Long Way to Go for Data Protection and Privacy Regulations
James Haillot Institute of Advanced Legal Studies (IHEJ-Paris)	Artificial Intelligence and Crime Prevention: Friend or Foe?
Lucia Martinelli Muse-Science Museum(Italy) Jiwon Shim Humanities Research Institue(Korea) Sungeu Chae Hanyang University(Korea)	Deep learning as boundary tools: perpetrating and detecting bias. The Case of COVID-19.
Mesut Serdar Çekin Turkey-Germany University Istanbul(Turkey)	Regulation of Artificial Intelligence
Ugo Pagallo University of Turin(Italy)	The Future of AI and the Society of the Future: Law, Ethics, and the Governance of AI

KEYNOTE SPEECH

AI and empathy: living in perfect harmony? Sharing some hopes and doubts

Amir Muzur University of Rijeka, Faculty of Medicine, Croatia

As in important feature of human intimate interactions, but also of some professional areas like healthcare or commerce, empathy is generally fostered and praised. No wonder, thus, that the obvious and well-documented decline in empathy, characteristic for modern societies, has stirred up the search for its artificial replacement: robots have been used as receptionists in hospitals, accompanying chatmates, and shop attendants.

By venturing into the cognitive and affective nature of empathy, this presentation tries to forsee some new possible applications of robotic artificial empathy as well as a few risks related to the narrowing-down of the colourfulness of interaction spectrum we are used to.

SESSION 1 : Human Behind Mechanical Intelligence

UNESCO'S work in the Field of Artificial Intelligence and the Concept of Transhumanism

Christian Byk Intergovernmental Bioethics Committee, UNESCO(France)

Julian Huxley, founder and first Director-General of UNESCO, is at the heart of contemporary debates on the nature and objectives of the concept of transhumanism, which he first used in the early 1950s. Therefore, the analysis of his idea of transhumanism should lead us to question his legacy in terms of the philosophy that inspires UNESCO's action to elaborate a comprehensive approach to artificial intelligence which aims to take into account the values and principles of universal ethics and aims to derive the best from the use of this technology

SESSION 1 : Human Behind Mechanical Intelligence

The Reach and Limitations of Machine Intelligence

Ivan Štajduhar Rijeka University(Croatia)

Creating machines that think and act like human beings has puzzled scholars from the dawn of time. While pondering on this concept, ancient Greek philosophers have established the rules that govern correct thought, consequently laying out the foundation for contemporary theorem provers and deduction systems. For quite some time, we believed that computation could be used to mimic reasoning, which in turn would lead to understanding the processes influencing behavioural patterns – it turned out, however, that it is difficult to encode rational thought. On the other hand, mimicking the thought processes going on in our brains, by reverse engineering of the brain, has proven to be infeasible, albeit it led to advances in some other fields of research. Regardless of the approach used for modelling the thought process behind decision making, working on machines that act like people was more or less abandoned because performing well in the imitation game (e.g. the Touring test) did not help in understanding human intelligence. Noticeable advances in artificial intelligence (AI) were reported only when the core focus shifted towards rational acting, disregarding the gist of previous approaches.

Nowadays, AI deals with the concept of creating machines thinking and acting like human beings in a rational sense, i.e. agents behaving optimally. Optimal behaviour can be learned (taught) using state-space search algorithms and self-play. Whereas rational acting by an autonomous agent can be considered solved for smaller scale problems in a simulated environment (games and such), the same cannot be stated for large scale problems, those involving uncertainty in a dynamic, ever-changing, environment (i.e., the real world). Here, an agent's performance largely depends on its ability to learn quickly, and from fewer examples – which can be helped by embedding the agents' percepts with machine learning (abstraction of the state space) and enhancing their search strategies using q-learning.

Recent advances in machine learning provided an end-to-end modelling framework for learning adequate feature embeddings, via stacked representations, directly from data (i.e. deep learning), which has proven to be rather useful for dealing with highly-nonlinear problems (e.g. those related to sound or vision). This, in turn, resulted in significant improvements in modelling numerous complex problems, previously considered infeasible for real-world applications (because of low fidelity), e.g. image to image translation, translation between written or spoken languages, and image inpainting, just to name a few. The same technology also triggered the development of techniques mimicking human abilities or appearance, e.g. artistic style transfer, speech synthesis, and so on. Although the aforementioned technologies can be utilised for building autonomous agents surpassing human experts in performing numerous highly specialised menial tasks, their intelligence (optimal rationality) is strictly limited to the task at hand, and they can often be easily fooled into suboptimal acting.

SESSION 1 : Human Behind Mechanical Intelligence

Breakfast with Plato, Dinner with Queen Min: Linguistic Avatars of GPT-3

Mariusz Pisarski University of Ss. Cyril and Methodius in Trnava(Slovakia)

Strikingly anthropomorphic effects of GPT-3, a third generation of AI language model developed by OpenAI, are bound to leave a lasting impact on public discourse, communication and education. This talk investigates conversational aspects of GPT-3, specifically the AI's ability to construct believable, linguistic representations of real people: historical or contemporary. As opposed to currently available chatbots and previous models of natural processing tools, the GPT-3 conversational simulacra seem to possess a form of intentionality (a will to be understood); are able to keep track of topics previously discussed; to correct themselves within a single conversation and even simulate a need of help and empathy. Engaging conversations with Plato, Queen Min or Princes Diana are now more possible than ever! Eerily, this can also apply to our deceased relatives. The only condition for GTP-3 is to have a trace such person leaves behind in vast corpora of the Internet.

Social, philosophical and ethical implications of the new generation of conversational simulacra are discussed. A proposed framework situates them within a broad historical tradition of agency transfers: from Eastern mythology (Shiva incarnating in the mortal realm) through computer game culture (avatars of players in the game world) to the post-human visions of transcending the body. The category of "avatarism" is introduced to describe various dynamics of avatar transfers and modes of interaction they afford. Although it is clear that when we send query text to GPT-3, we are not communing with a unique "digital soul", a current studies suggest possibility of emotional attachment to simulated entities. If so, then psychological, ethical or even legal issues need to be discussed, even if one considers such enquiries as limited to Wittgenstein's closed circle of language.

By treating the anthropomorphic simulacra as avatars and demonstrating a continuity between ancient mythologies and 21st century AI technologies a vast reservoir of cultural resources is revealed to serve as a reference toolkit for future studies of the effects such interaction could have on social discourse. The focus of attention in this paper not on cognitive or computational aspects of the new AI language model, but on a possible impact of its mimetic representations on existing notions of identity, agency, immortality. As such, it hopes to contribute to bridging the gap between discourses and bringing AI to Humanities, and Humanities to AI.

BIBLIOGRAPHY

Askell, Amanda (2020). GPT-3: Towards Renaissance Models, in: "GPT-3 and General Intelligence", in: Philosophers On GPT-3 (updated with replies by GPT-3), online: https://dailynous.com/2020/07/30/philosophers-gpt-3/#zimmermann
Barfield, Woodraw (2015). Cyber-Humans. Our Future with Machines, Springer.
Baldwin, Sandy (2015). The Internet Unconscious: on the Subject of Electronic Literature, Bloomsbury.
Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press.
Braidotti, Rosi (2013). The Posthuman. Polity, 2013.
Chalmers, David (2020). GPT-3 and General Intelligence, in: Philosophers On GPT-3 (updated with replies by GPT-3), online: https://dailynous.com/2020/07/30/philosophers-gpt-3/#zimmermann
Gwern Branwen. GPT-3 Creative Fiction, online blog: https://www.gwern.net/GPT-3
Kaiya Case (2019). GPT-2 AI Poetry Generation: Writing like Donne, IPHS 200.0l Programming Humanity, online: https://digital.kenyon.edu/dh_iphs_prog/17/
Müller, V. C. (Ed.) (2016). Fundamental Issues of Artificial Intelligence, Synthese Library; Berlin: Springer.
Rini, Regina (2020). The Digital Zeitgeist Ponders Our Obsolescence in: Philosophers On GPT-3 (updated with replies by GPT-3), online: https://dailynous.com/2020/07/30/philosophers-gpt-3/#zimmermann
Sautoy, Marcus (2020). The Creativity Code. Harvard University Press.
Shevlin, Henry (2020), A Digital Remix of Humanity, in: Philosophers On GPT-3 (updated with replies by GPT-3), online: https://dailynous.com/2020/07/30/philosophers-gpt-3/#zimmermann
Thacker, Eugene (2010). After Life. University of Chicago Press.

SESSION 1 : Human Behind Mechanical Intelligence

Second wave AI and new cognitive technologies: the question of transparency and trust

Olga Markic University of Ljubljana(Slovenia)

New cognitive technologies and AI are nowadays used as tools by the experts and in different areas of everyday life. However, despite massive usage of such tools, people often lack proper knowledge of how they actually work. For example, deep neural networks models are designed to solve various machine learning task in a way that is potentially opaque for humans. This concerns both the producers and the users of the tools. In this paper I will briefly describe two waves of AI. The first wave (also called GOFAI) is based on propositional knowledge and deduction, and is using symbol manipulating approach to modeling. In contrast, the second wave is based on non-conceptual knowledge, induction and learning. It is using machine learning and neural network approach to modeling. The latter approach is using statistical pattern matching over large data sets. These data sets that computer scientists are using for learning processes play an important role in the behavior of the model. In the second part of the paper I will examine the examples of second wave approach, focus on the question of transparency and connect it to the question of trust. It is now known that models may be inherently biased. Scientists have also found out that there are vulnerabilities in machine learning systems which allow a small, carefully designed change in how inputs are presented to a system, to completely alter its output. These adversarial attacks cause manifestly wrong conclusions that may be intentionally crafted but very hard to find out. I will argue that despite great successes of the second wave in, for example, solving problems with face and voice recognition, with recognition of intentions and with designing programs for playing games (e.g. Alpha Go), philosophical reflection on the ontological and epistemological assumptions of the approach is needed. We have learned about the inadequacy of the GOFAI, and we should be cautious of the preliminary success of the second wave.

SESSION 1 : Human Behind Mechanical Intelligence

How Young Minds See Future Minds: Uncanny Valley and Ascription of Different Types of Attributes to Robots

Robin Kopecký Karel Capek Center, Charles University(Czech Republic)

The question of the impact of new technologies on human society becomes ever more pressing. Children of next generations will be born into the world full of new stimuli that will form their thinking in the crucial years of development. Nowadays interactions between child and machines are being studied on both philosophical and technical levels (Tung 2011, Belpaeme et. al 2013). It is possible that children will ascribe certain specifically human traits to robots that will exhibit certain behavioural patterns. In our interview study with children and teenagers (N=209, 109 F, age range 6-17, mean age=11.05) we decided to test the children's reaction to different types of robots. In the first part of the interview, the children were shown pictures of 6 different robot faces created by an industrial designer. The pictures were intended to be a spectrum from the most robotic to the most human-like face. The children were supposed to judge how much the particular face appears friendly to them on a 6-point Likert scale. As we hypothesised, in the resulting graph we observed an obvious uncanny valley. The friendliest robot was the one with eyes, but without other facial features, while the least friendly was the robot with human facial features set on the metal head, together with two robot heads with no facial features whatsoever. The robot that had all the facial features, hair and skin, scored as the second most friendly-looking robot. In the second part of the interview, the children were asked to judge whether they would ascribe certain attributes (sense perception, thinking, emotions, self-reflexion, life, soul and freedom) to different beings or entities described to them. The training examples were a dog and a child. Then they were asked about a robot with basic perception and motor abilities that can also do complicated maths and talk, and a robot who is also capable of being an active participant of a dialogue and can talk about how it perceives the world in a way indistinguishable from a real person. The results showed that children were more prone to ascribe certain human abilities to the android than to the mathematical robot (esp. emotions and freedom). Both robots scored highly in sense perception and thinking, and android scored also in self-reflexion. However, even the android didn't score when it comes to the ascription of life and the soul. Participants ascribed soul and life only to the living beings.

The results show that the children are capable to ascribe certain level of cognition, emotions and even freedom to sufficiently complex robots, but are still reluctant to see robots as living beings with a soul. The concept of the soul may be understood as an essential moral core of the person that is capable of moral conduct and deep interpersonal bonds and which is conceptualized differently than the mind (Richert & Harris 2006, 2008). Despite their behavioural complexity and intelligence, robots might not manage to cross the boundary between machines and full-fledged persons in the eyes of children.

SESSION 1 : Human Behind Mechanical Intelligence

The Emotive Machine, and Mechanised Self

Aditi Vashistha University of Delhi(India)

The self exists in interaction with the other selves. All of them shape the creation of subjectivity. With time, machines are created taking into account the human emotions. Humans have adopted according to the machines. The human capacities are reframed and changed. From the health to the emotional states.

There are machines that are mediators to not only the world outside but also to our own body. There are apps to give account of the health and to know emotional state of people. This mediation has also changed many living experiences such as feeling like love, friendship, attention and support. The social concept such as fidelity and infidelity. For example, the 2013 movie Her by Spike Jonze, where the love story of a computer programme and the human is shown. The concept of intimacy and relationship are reframed. In the dystopic tv series, 'Black Mirror's episode 'Striking Vipers'. There are two male friends who play a video game. Where one plays as woman, and the other as a man. They develop a sexual relationship. This kind of possibilities are creating new complexities in terms of emotions.

Machines are not only becoming emotional. They are creating new emotionally challenging situations. The machines are not only a mediator. But they are active creators. They are shaping a completely different life form. The self in the virtual world is different from the self in real world. It is not only the real that shape virtual or impact it. But at the same time, the virtual impacts the real. For example, for an Instagram influencer, it is the virtual self that constantly impact or intervene in the real self.

The purpose of this paper is to read machines as active agents rather than non-living things and read humans as machines, constantly shaped by the tools they are formed with i.e. their environment. For this purpose, the paper reads the broader outline of science fiction genre in movies and fiction. To see this transformation, it will be taking into account the Philip K. Dick's book 'Do Androids Dream of Electric Sheep? (1968) And its adaption in the film Blade Runner 2049. The idea of human and the replicant that is a bioengineered being (who looks identical to humans but have superior strength, speed and intelligence) is an interesting exploration into the relationship between mechanised human self and emotive machines.

BIBILOGRAPHY

2011. Black Mirror, Season 5.
Deleuze, Gilles. 1988. Spinoza: Practical Philosophy . City Lights Book.
Dick, Philip K. 1999. Do Androids Dream of Electric Sheep? . Orion Books.
Guattari, Gilles Deleuze and Felix. 1977. Anti- Oedipus: Capitalism and Schizophrenia. The Viking Press.
2013. Her. Directed by Spike Jonze. Performed by Scarlett Johansson Joaquin Phoenix.
2017. Blade Runner 2049. Directed by Denis Villeneuve. Performed by Harrison Ford Ryan Gosling.

SESSION 1 : Human Behind Mechanical Intelligence

Heidegger's Tools and Technology Criticism & the Era of Artificial Intelligence

Jiseon Park Hankuk University of Foreign Studies(Korea)

Heidegger is a German existentialist philosopher who represents 20st-entury modern philosophy. He went through both World War I and II, and witnessed technological civilization killing humans in large quantities. Against this background, Heidegger was very critical of the modern epistemological tradition in which the subject created by modern philosophy was immersed in how to perceive objects. Modern epistemology has contributed to man's escape from God's oppression, conquering nature and developing capitalist technical civilization. However, on the other hand, it has resulted in the negative consequences of the loss of humanity in which the subject oppresses and controls others. From this point of view, Heidegger proposed the question of human existence, trying to replace the philosophical agenda that Western modern philosophy has focused on developing epistemology.

Heidegger presented a tool (Zeug) as a critical role concept in his philosophical system. In Heidegger's philosophy, tools appear to Dasein (human) through surface function and are used for any purpose. Here, tools mean the state used by users to do something naturally, associating with other tools in this world. Heidegger called the tool's existence 'Zuhandenheit (readiness-to-hand)' which can be interpreted directly as 'in the hand'. In other words, in Heidegger's philosophy, it is considered that Dasein (human) must have tools in hand to escape from its own existence as being-in-the-world, and these tools and the Dasein must be fused familiarly with each other.

Meanwhile, Heidegger can be understood as the first classical philosopher to directly criticize modern science and technology. He considered that in the modern age of science and technology, it was only a part of the human capitalist civilization, and this capitalism drove humans into parts.

Using this concept of Heidegger, this study aims to critically consider the current technological development trends of artificial intelligence and the future of the artificial intelligence era. Deep learning technology, which currently leads artificial intelligence technology, is showing very good performance in image recognition and is putting to practical use. Image recognition technology centered on deep learning has a very similar character to the epistemology of modern Western philosophy. Artificial intelligence is expected to play a role in recognizing things and controlling human beings and civilizations. Therefore, this study will criticize the "'Zuhandenheit (readiness-to-hand)" suggested by Heidegger and the role of artificial intelligence, critically suggesting similarities between Western recognition theory and deep learning technology. Besides, this study will warn against the danger of artificial intelligence pushing humans away, based on Heidegger's philosophy that criticizes humans being driven into parts of technological civilization.

하이데거는 20세기 현대 철학을 대표하는 독일의 실존주의 철학자이다. 그는 양차 대전을 모두 겪으며 기술 문명이 인간을 대량으로 살상하는 것을 목격하였다. 이러한 배경에서 하이데거는 근대 철학이 창조한 주체가 객체를 어떻게 인식하는 것에 몰두한 근대 인식론적 전통에 매우 비판적이었다. 근대 인식론은 인간이 신의 억압에서 벗어나 자연을 정복하고 자본주의 기술문명을 발전시키는데 기여하였지만, 다른 한편 주체가 타자를 억압하고 통제하는 인간성 상실의 부정적 결과를 야기했다. 이러한 관점에서 하이데거는 인간의 존재의 문제를 제시하면서 서양 근대 철학이 인식론을 중심을 발전시켜온 철학적 아젠다를 대체하려고 하였다.

하이데거는 그의 철학체계에서 핵심적 역할 개념으로 도구(Zeug)를 제시하였다. 하이데거의 철학 속에 도구는 현존재 (인간)에게 표면적 기능을 통하여 나타나서 어떠한 용도로 사용되는 존재이다. 여기서 도구는 이 세상 안에서 다른 도구들과 연관하면서, 사용자에 의해 자연스럽게 어떤 일을 하기 위해서 사용되는 상태를 의미한다. 하이데거는 이러한 도구의 존재방식을 직역으로 '손안에 있음'으로 해석할 수 있는 '용재성 (Zuhandenheit)'이라고 명명하였다. 즉, 하이데거 철학에서 현존재(인간)은 세계내존재로서 자신의 존재로부터 탈피하려면 도구가 손안에 있어야 하고 이러한 도구와 현존재가 서로 친숙하게 융합 되어야 한다고 보았다.

한편 하이데거는 현대 과학기술에 대하여 직접적으로 비판을 한 최초의 고전적 철학자로 볼 수 있다. 그는 현대 과학기술 시대에는 인간자본주의 문명 속의 부품에 불과하고 이러한 자본주의는 인간을 부품으로 내몰아간다고 보았다.

본 연구자는 이러한 하이데거의 개념을 사용하여 현재 인공지능의 기술적 발전 동향과 인공지능 시대의 미래를 비판적으로 고찰하고자 한다. 현재 인공지능 기술을 주도하도 있는 딥러닝 기술은 이미지 인식에서 매우 뛰어난 성능을 보여주고 있고 빠른 속도로 실용화되고 있다. 딥러닝을 중심으로 한 이미지 인식기술은 근대 서구 철학의 인식론과 매우 유사한 성격을 가지고 있다. 인공지능이 사물을 인식하여 인간을 지배하고 문명을 통제할 수 있는 역할을 할 것으로 예측된다. 따라서 본 연구자는 서양인식론과 딥러닝 기술의 유사성을 비판적으로 제시하면서 하이데거가 제시한 도구의 '용재성'과 인공지능의 역할을 비판할 것이다. 아울러 인간이 기술문명의 부품으로 몰려짐을 비판한 하이데거 철학에 착안하여 인공지능이 인간을 내몰아가는 위험성을 경고할 예정이다.

SESSION 1 : Human Behind Mechanical Intelligence

Human dignity and surveillance society in the age of artificial intelligence: centering on Jeong, je-du, Wang, Yang-Ming and H. Begson, M. Merleau-Ponty

Seonjin Yang Chung-Ang University(Korea)

Modern science and technology are atomic and causal ontology or worldview that separ ate human from nature and regard nature as a machine. However, the so-called 4th Indu strial Society, a new era, may have to revise such an ontology or worldview. Because 4th i ndustrial society will lead to a "connected society" that brings together humans and hu mans, humans and things, humans and nature. In the era of the 4th Industrial Revolution , the cultural and spatial characteristics defining human beings will gradually disappear a nd enter into a new relationship based on digitalization. This relationship may lead to a n ew explanation of human identity.

The new digital society is fundamentally changing the relationship between human-hu man, and human-society, and the era will come when we need to existentially recognize the relationship between humans and things which are excluded from the existing world view or theory of relationships. Responding to the changes in the new era, this study aim s to consider Wang Yangmin of the East and Merleau-Ponty of France. Because they are p hilosophers who deny atomic and mechanistic existences or worldviews and present ont ology or worldview in which human-nature, human-human, and human-thing are integr ated. To build a new ontology in the new era, this study looks at the theories of the traditi onal philosophers, that is Wang Yang-min and Merleau-Ponty, to find clues to an ontolog y appropriate to the era of the 4th Industrial Revolution.

근대의 과학기술은 인간과 자연을 분리하고 자연을 마치 기계처럼 간주하는 원자론적이며 인과론적 존재론 내지 세계관이다. 하지만 새로운 시대인 일명 제4차 산업 사회는 기존의 존재론을 수정해야할지 모른다. 왜냐하면 제4차 산업 사회는 인간과 인간, 인간과 사물, 인간과 자연을 하나로 묶는 '연결 사회'를 초래할 것이기 때문이다. 4차 산업혁명 시대에는 인간을 규정하는 문화적 특성과 공간적 특성은 점차 사라지고 디지털화에 기반을 둔 새로운 관계 속에 진입하게 될 것이며 이러한 관계는 새롭게 인간의 정체성을 설명해야하는 상황이 발생할 수도 있을 것이다.
새로운 디지털 사회는 인간과 인간, 인간과 사회의 관계를 근본적으로 변화시키고 있으며 기존의 세계관 내지 관계론에서 배제되었던 영역인 인간과 사물의 관계를 존재론적으로 인정해야하는 시대가 다가올 것이다. 논자는 새로운 시대적 변화에 부응하는 철학자로서 동양의 왕양명과 프랑스의 메를로-뽕띠를 통해서 고찰하려고 하였다. 왜냐하면 두 철학자는 원자론적이며 기계론적인 세계관 내지 존재론을 부정하고 인간과 자연, 인간과 인간, 인간과 사물을 하나로 묶는 존재론 내지 세계관을 제시한 철학자들이기 때문이다. 새로운 시대에 새로운 존재론을 구축하기 위해서 우리는 전통 철학자인 왕양명과 메를로-뽕띠의 존재론을 살펴보고 제 4차 산업혁명의 시대에 부응하는 존재론의 실마리를 찾고자 하였다.

SESSION 2 : The Promise and Perils in the Era of Technology

Using the Kübler-Ross Model and a Proactive Approach Toward Managing Digital Fear by Adopting Iterative and Participatory Strategies on a Research-Literacy-Regulatory Spectrum

Kuei Tien Chou · Ngerng Yi-Ling National Taiwan University Risk Society and Policy Research Center (Taiwan)

The advent of new digital technologies such as artificial intelligence (AI) and their increasing use has led to fears and privacy concerns over how the technology is being used. However, if we were to understand that fear of technology is a natural reaction to each rise of a new technology, it is possible to adopt proactive and iterative strategies to mediate the impact of such fears, and to find a balance between technological innovation and technology concerns. The Kübler-Ross Model is a framework by which to guide these strategies.

Fear of new technology and corresponding privacy concerns have existed for as long as new technology arose (Lepore, 2013). Studies across various countries have found the older generation to be more concerned about privacy as compared to the younger generation, while the younger generation also have greater knowledge of privacy settings on social media platforms such as Facebook, and are also more likely to adopt various strategies to protect their privacy (Pingitore et al., 2013; Van den Broeck et al., 2015; Blank et al., 2014; Kan, 2019).

The Kondratieff cycles illustrate that the global economy has undergone economic fluctuations of 40 to 60 years per cycle, with the start of each cycle coinciding with the advent of new technological innovations, and as the technology use becomes more widespread and integrated into the economy, the cycle comes down from its peak before another new cycle develops with the rise of another new technology. Five cycles have been identified via the development of the steam engine in the late 1700s, the railway and steel industries in the early 1800s, electrification in the late 1800s, automobiles and petrochemicals in the early-1900s, and the information and communications technology in the late 1900s. Intelligent technologies like AI could give rise to a new cycle (Wilenius and Kirki, 2014). The fear and privacy concerns that follows each new technological innovation can be explained by the “cultural lag” concept which highlights that non-material culture such as values and beliefs tend to lag behind the development of material culture such as new technology (Ogburn, 1966; Marshall, 1999).

This article proposes that the Kübler-Ross model developed initially to help individuals cope with grief could be used as a starting point to help understand digital fear (Kübler-Ross et al., 1972). In the context of digital fear, individuals who experience a new technology for the first time could in the first stage enter into 'shock' and 'deny' the use of the technology in certain aspects of their life (for example, due to data privacy concerns on Facebook), before the increasing encroachment of the technology results in 'anger' (the next stage) and they advocate opposing the use of the technology (as illustrated by the Cambridge Analytica incident). As a new technology becomes more prevalent and moves toward greater understanding, individuals may learn to 'bargain' over the appropriate use of the technology (such as the level of privacy to allow for Facebook's AI algorithms to obtain personal data), before they enter into a stage of 'depression' where they come to the realization that their rights might still be infringed upon as use of the new technology evolves, and then eventually coming to an 'acceptance' of the new technology (to stop using Facebook entirely or to accept the privacy options Facebook offers). Using the Kübler-Ross change curve to understand how individuals cope when they encounter a new technology could enable strategies to be developed to address their concerns early-on and to encourage greater understanding of the technology and awareness of its potential dangers, as well as mediate its effects, so as to balance the innovation impetus alongside digital concerns.

Two foresight workshops conducted by the Risk Society and Policy Research Center at the National Taiwan University were conducted with 17 AI, data and digital experts in Taiwan to further understand their data concerns and proposed strategies for mediation, where the following insights and recommendations were obtained. The findings were matched to each stage on the Kübler-Ross model. To understand the shock and fear that initially emerges due to digital fear, social research should be conducted alongside new technological innovation to understand these concerns. Digital literacy programs should correspondingly be developed for each stage sector, including among innovators and the government (to enable greater understanding of digital concerns), so as to reduce the possibility that technology misdevelopment and misuse could lead to anger over its use. As new technologies bring with it a steady stream of new concerns during its initial stages, participatory citizen engagement should be conducted on an iterative basis to continually understand the new fears that may arise with new understandings of the technology, and to develop relevant digital literacy tools address these fears. Regulations should also be developed and continually updated with input via these participatory workshops, as well as from the independent oversight bodies that need to be set up to regulate the use of the new technology. In the case of AI, an independent digital and data committee and a consumer data protection body for complaints could be set up.

The process by which oversight should be provided over each new technology should therefore be a proactive process where digital social research, literacy and regulation should be viewed as existing on a spectrum of strategies, and conducted alongside technological research and innovation, where the development and implementation of these strategies should be done on an iterative basis so as to continually address the concerns that would arise with the adoption of a new technology at each stage, and to thereby create a coping process by which the introduction of a new technology can be managed, to both encourage technological innovation while addressing the needs of individuals. By developing a systemic framework to cope with the rise of each new technology, such proactive intervention would enable programs and policies to be enacted at an earlier stage to deal with the negative side-effects of new technology.

BIBILOGRAPHY

Lepore, J. (2013). The Prism: Privacy in an age of publicity. *The New Yorker*, (June 24, 2013). Retrieved from: https://www.newyorker.com/magazine/2013/06/24/the-prism

Pingitore, G., Meyers, J., Clancy, M., & Cavallaro, K. (2013). Consumer concerns about data privacy rising: What can business do. *Retrieved October*, *29*, 2014. Retrieved from: https://fpf.org/wp-content/uploads/JDPA-Data-Privacy-Research-Oct-2013.pdf

Van den Broeck, E., Poels, K., & Walrave, M. (2015). Older and wiser? Facebook use, privacy concern, and privacy protection in the life stages of emerging, young, and middle adulthood. *Social Media+ Society*, *1*(2), 2056305115616149. Retrieved from: https://journals.sagepub.com/doi/full/10.1177/2056305115616149

Blank, G., Bolsover, G., & Dubois, E. (2014, August). A new privacy paradox: Young people and privacy on social network sites. In *Prepared for the Annual Meeting of the American Sociological Association* (Vol. 17). Retrieved from: https://www.oxfordmartin.ox.ac.uk/downloads/A%20New%20Privacy%20Paradox%20April%202014.pdf

Rainie, L. (2016). The state of privacy in post-Snowden America. *Pew research center*, *21*. Retrieved from: https://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/

Kan, M. (2019). Mozilla releases the Taiwan Internet health survey report – the top three issues to be on alert for are privacy risks, personal data collection and fake news. Retrieved from: https://www.mobile01.com/topicdetail.php?f=508&t=5888166

Wilenius, M., & Kirki, S. (2014, November). Riding the Sixth Wave: Kondratieff Theory as a Method in a Multi-Stakeholder Process for the Renewal of the Finnish Forest and Financial Services Industries. In *5th International Conference on Future Oriented Technology Analysis: Engage Today to Shape Tomorrow, Brussels*. Retrieved from: https://ec.europa.eu/jrc/sites/jrcsh/files/fta2014_presentation_t2s_130_wilenius.pdf

Ogburn, W. F. (1966). *Social change with respect to cultural and original nature* (No. HM101 O4 1966).

Marshall, K. P. (1999). Has technology introduced new ethical problems?. *Journal of business ethics*, *19*(1), 81-90. Retrieved from: https://link.springer.com/content/pdf/10.1023%2FA%3A1006154023743.pdf

Kübler-Ross, E., Wessler, S., & Avioli, L. V. (1972). On death and dying. *Jama*, *221*(2), 174-179.

SESSION 2 : The Promise and Perils in the Era of Technology

Human Work in the Digital Economy

Tomáš Hříbek Filosofický ústav AV ČR/Institute of Philosophy(Czech Republic)

It is generally regarded as highly plausible that AI and robots will eventually take over many, or most, occupations so far carried out by humans. If this indeed happens, it will cause what might be called the Pauperization Problem. Large numbers of people will lose their livelihood. I provide a brief critical examination of three attempted solutions to the Pauperization Problem: (1) Universal Basic Income, (2) Collaborative Commons, and (3) the Social-Digital Economy. I find each of them lacking in some respects. And yet, I consider the Pauperization Problem largely technical. There is a deeper philosophical issue concerning the transition to the post-work world, i.e. finding a meaning and flourishing in such a world.

SESSION 2 : The Promise and Perils in the Era of Technology

Digitocracy: Ruling and Being Ruled

Alfonso Ballesteros Universidad Miguel Hernández de Elche(Spain)

A significant amount of research in recent years has focused on digitalisation: On its new forms of power, and on how it is shaping society and man. Frank Pasquale gave an in-detail analysis of the first two elements (power and society) and the technologies used to control information and money with algorithms (from Silicon Valley and Wall Street). Éric Sadin has shown the stages and spirit behind Silicon Valley's rule. Byung-Chul Han argues that the peculiar efficiency of this power through digital technology resides in the exploitation of pretended individual choices. It is the so-called psychopolitics, the predominant form of power in post-industrial capitalism. The work of the Center for Humane Technology is also essential to understand this new form of government and the resulting man. This centre is the most crucial think tank on digitalisation, based in Silicon Valley, although it acts as a kind of conscience of the Valley.

Many scholars have studied digitalisation as a process that will lead us to an inhuman world through strong artificial intelligence (AI). This perspective seems to be problematic, as it takes digitalisation as an unavoidable necessity. Moreover, it forgets to question current digitalisation and how it shapes power, society, and man. In a similar vein, other scholars take for granted that we live in an "AI society" and that digitalisation is something that needs to be moralized with principles. This might be misleading. On one hand, we cannot take anything for granted—we do not live in an AI society. Moreover, principles might be used to avoid discussion because we only formulate principles relevant to those things we accept. Of course, many documents with principles to regulate AI have appeared, and these are useful and necessary. However, they are not enough. We need further discussion on this topic, both by scholars and by laypersons. Digitalisation is a political matter. For these reasons, more work is needed to describe and evaluate digitalisation.

The purpose of this study is to describe and examine digitalisation as a form of government: Digitocracy. This work investigates the relationship between those who rule and those who are ruled; in particular, how the latter are shaped by digitalisation. Therefore, from it emerges a new kind of man: Animal digitalis. This presentation is inspired by Hannah Arendt's political philosophy, whose humanism allows us to show the superiority of humans over animals and artefacts—precisely, a distinction that digitalisation is blurring.

The main thesis of this presentation is that digital technology as it is today tends to animalise man. This is the main novelty of this work. The thesis is built in the philosophy of Hannah Arendt and Byung-Chul Han. The former considered man in the 20th century animalised, but described it then as machine-like in terms of language, as she did not live during the irruption of digital technology and its animalisation of language with emotion. The complement to Arendt is Byung-Chul Han, who explicitly reasons against Arendt's assumption of animalisation; however, I argue that his philosophy confirms hers, and also serves to describe the animalisation of 21st century man by digital technology. As digitalisation expands even more as the Covid-19 crisis also expands, our understanding of it is even more necessary.

SESSION 2 : The Promise and Perils in the Era of Technology

Fighting the Machine or to Fight Capitalism?

Avantika Tewari Jawaharlal Nehru University(India)

As the pandemic rages on, the technologies to trace, test and track the body have become of central importance ringing alarm bells in people and rightfully so, considering the surveillance capitalist advancements. Indeed, there is a shift even to that, in the form of the development of what has been called, “bioinformation, and the bioinformational paradigm where these two forces of new biology and information coalesce, overlap, and intermingle in the logic that drives bioinformatics and bioinformational capitalism that is self-renewing in the sense that it can change and renew the material basis for life and capital as well as program itself.”1) There has been a temptation to look at the penetration of technology in society as an inauguration of a techno-authoritarian regime. Agamben says, “What becomes of human relations in a country that grows accustomed to living in this way for who one knows how long? What is a society that cleaves to no value other than that of survival?”2) A similar refrain came from Henry Kissinger who describes the catastrophic consequences may not only be physical but also apply to humans’ conceptions of themselves.3)

The question of the machine-human is posited in exceptional and oppositional terms to shrink and shape the conversations around them, what is often occluded by such a framing is the question of social organization of resources, means of production and ownership of those means. With life increasingly being made profitable data points, it is crucial how we ground our resistance. Imposing a universal discourse of life, which insists on ethics, tactics, democratic universal principles, morality – as much as they are crucial, is an ideal abstraction of society which underscores the mediations of social relations. Every epoch has seen technological advancement; some even at the scale of acquiring ‘singularity,’ in fact the hybrid experiments we see today have semblances to the kind of experiments that were taking place in 1920s USSR.4)

People across the political spectrum are invested in emphasizing the COVID-19 pandemic as a logical outcome of human transgressions or systemic excesses thus revealing the moral depravity and predatory 5) nature of capitalism and yet they find reconciliation in the existing modes of social organization of life and labour. Such solution-oriented conversation around the 'practical' response to a mighty surveillance regime prevents us from discussing the matter, politically. While I share Agamben's concern for bio-medicinal colonization of life, I do so without suffering from any nostalgia for non-mechanized life. Rather, I content that; biometric incursions by capital changes very little of the world except in the form of its appearance; thwarting creativity and innovation while speaking its language. In fact, what this moment reveals to us is how the excesses that we attribute to capitalism, reveals it truth.

"Machinery is no more an economic category than the bullock that drags the plough. Machinery is merely a productive force. The modern workshop, which depends on the application of machinery, is a social production relation, an economic category...There is nothing more absurd than to see in machinery the antithesis of the division of labour, the synthesis restoring unity to divided labour. The machine is a unification of the instruments of labour and by no means a combination of different operations for the worker himself." 6) (Marx 116, Poverty of Philosophy)

Therefore grounding the basis of our struggle on the grounds on privacy and sacredness of moral life cannot get us too far. To suggest that capitalists have no business invading our privacy is to ignore the fact that they are in the business of profiteering off of life, turning our everyday activities and digital footprint into surplus value. This is suggestive of the very vampirism of capital and not only the authoritarianism of the State which is seen as the one passing policies that disable citizen's privacy in favour of corporate interests.

Tony Smith observes, 7) "Technology is more than just a weapon for inter-capitalist competition; it is a weapon in struggles between capital and labour. Technological changes that create unemployment, de-skill the workforce, and enable one sector of the workforce to be played against another shift the balance of power in capital's favor." In such times, the question of the machine and human relations is as much about the social organization of relations between them, therefore opening a window to go beyond capitalism as much as it poses a risk to regress into authoritarianism. How do we approach the question of equality under such times that is not confined to increasing the superficial boundaries of privacy as a defense against techno-capitalists?

BIBILOGRAPHY

1) Michael A. Peters, "Philosophy and Pandemic in the Postdigital Era: Foucault, Agamben, Zizek" April 29, 2020. Accessed at https://link.springer.com/article/10.1007/s42438-020-00117-4

2) Giorgio Agamben, "Clarifications" tn. Adam Kotsko, March 17, 2020: https://bit.ly/2EpZCsH (accessed August 11, 2020)

3) Henry Kissinger, "How the Enlightenment Ends." *The Atlantic*, May 15, 2018. Accessed at https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger -ai-could-mean-the-end-of-human-history/559124/.

4) See Nicholas D. Wright, '*Artificial Intelligence, China, Russia and the Global Order*,' Air University Press, Alabama: 2019

5) See "Will Fear of COVID-19 and Loathing of Capitalism Drive Systemic Change?" Apr 22, 2020, Accessed at https://www.csrwire.com/press_releases/44816-Will-Fear-of-COVID-19-and-Loathing-of-Capitalism-Drive-Systemic-Change-

6) Karl Marx, '*Poverty of Philosophy,'* Originally Published in 1847, Re. Progress Publishers, 1955, Tn. Institute of Marxism Leninism, 1955

7) Tony Smith, 'How Capitalism Stifles Innovation," *Tribune*, August 30, 2020.

SESSION 2 : The Promise and Perils in the Era of Technology

Can Artificial Intelligence Promote Consumption? Comparison of 'Say's Law and 'Demand Control'

Suhyoun Hwang Sogang University(Korea)

In economics, demand and supply are the most important and basic concepts. In classic al economic theory, supply and demand are priced by the 'invisible hand' in the market, and transactions are made freely. However, the economic depression of the 20th century and the world war raised fundamental questions about this laissez-faire economic theor y. According to the theory claimed by Say among the classical economic theory laws, wh en supply is natural, products are traded in the market without separate measures. In ot her words, if there is no problem with supply, there will be no economic problem. Suppli er-oriented economics was adopted by the Reagan administration in the 1980s, becomin g the foundation for neo-liberalism.

Meanwhile, British economist Keynes thought that laissez-faire market policies could fail and that the government should directly intervene in the market to supplement the mar ket economy. However, Keynes thought that supply could control demand. It means that the government could curb demand or create demand by intervening in the market. The se economists' approach to supply and demand was closely related to the political lands cape of Western economic bloc.

In this context, this study aims to conduct a comparative study on how the distribution in dustry utilizes artificial intelligence and data science and how macroeconomics respond s to the market economy. The distribution industry currently seeks to increase companie s' competitiveness by collecting large amounts of information from consumers and appl ying various methods of identifying consumers' patterns and behavioral psychology. It is particularly active in introducing data science and artificial intelligence while taking the l ead in responding to the recent spread of artificial intelligence and data industries.

This study considers that the way distribution companies use artificial intelligence is similar to Keynes' acceptance management. In this context, this study did analyze the use cases of artificial intelligence used in the distribution industry. In the future, the AI era is expected to see Keynes' demand management outstripping Say's supplier-centered approach. Through this analysis, the future political landscape is expected to be a favorable environment for converting to demand management rather than laissez-faire.

경제학에서는 수요와 공급이 가장 중요하고 기본적 개념이다. 고전적 경제학 이론에서는 수요와 공급이 시장에서 '보이지 않는 손'에 의하여 가격이 결정되고 거래가 자유롭게 이루어진다. 그러나 20세기 경제불황과 세계대전은 이러한 자유방임주의 경제이론에 대한 근본적 의문을 제기하였다. 고전적 경제 이론 법칙 중 세이가 주장한 이론 공급이 자연스럽게 이루어지면 별도의 조치가 필요없이 시장에서 제품이 거래가 이루어진다고 보았다. 공급의 문제가 발생하지 않으면 경제적 문제가 발생하지 않는다고 보았다. 공급자 중심의 경제학은 1980년대 레이건 정부에서 채택되어 신자유주의 기초적 토대가 되었다. 한편 영국의 경제학자 케인즈는 자유방임적 시장정책은 실패할 수 있고 정부가 시장에 직접개입해서 시장경제를 보완해야 한다고 보았다. 그런데 케인즈는 공급이 수요를 통제할 수 있다고 보았다. 정부가 시장에 개입하여 수요를 억제하거나 수요를 창출할 수 있다고 보았다. 이러한 경제학자들이 공급과 수요에 접근하는 태도는 서구 경제권의 정치지형과 밀접하게 연관되었다. 이러한 배경에서 본 연구자는 유통산업이 인공지능과 데이터 과학을 활용하는 방법과 거시경제학이 시장경제에 대응하는 태도를 비교연구를 수행하고자 한다. 현재 유통산업은 소비자의 대용량의 정보를 수집하여 소비자의 패턴과 행동심리를 파악하는 다양한 방법을 적용하여 기업의 경쟁력을 높이려고 한다. 특히 최근의 인공지능과 데이터 산업의 확산에 선도적으로 대응하면서데이터 과학과 인공지능의 도입에 적극적이다.

본 연구자는 유통기업이 인공지능을 활용하는 방식이 케인즈의 수용관리와 유사하다고 판단하고 있다. 이러한 배경에서 본 연구자는 유통업계에서 활용되는 인공지능의 활용사례를 분석하면서 향후 인공지능 시대는 케이즈의 수요관리가 세이의 공급자 중심의 방법보다 앞서 나갈 것으로 판단된다. 이러한 분석을 통하여 미래 정치지형도는 자유방임주의보다 수요관리 중심으로 변환되기 유리한 환경으로 예상한다.

SESSION 2 : The Promise and Perils in the Era of Technology

The Implications of ARTIFICAL INTELLIGENCE

Sanghyong Shin Andong National University(Korea)

We examine the context of using artificial intelligence in various ways. Ultimately, focusing on augmented reality, we analyze how artificial intelligence and human intelligence contact and communicate. Therefore, it is emphasized that in order to overcome this, it is necessary to clarify the cause of the opposing relationship between the two in the present state and go to a mutually beneficial relationship. In this respect, the position of human and human intelligence is modeled not as a relationship between instruction and execution, order and delivery, and the principal and the assistant, but as a complementary relationship in which each other cooperates with the goal of optimization. And it tries to explain that the ultimate goal of this collaboration is not ingenious conception and creation, but realization that includes the intermediate goal of better naturalization. In addition, I would like to emphasize that the focus of the artificial intelligence era is considering psychological diversity.

SESSION 2 : The Promise and Perils in the Era of Technology

A Study on the Values and Humanism of Confucian Society from the Social Engineering Perspectives of Artificial Intelligence

Yeonjae Kim Kongju Natinal University(Korea)

This study begins with a question of how to embrace artificial intelligence in modern society, paying attention to artificial intelligence discourse in the era of the 4th revolution. As a methodological idea of approaching this, this study aims to discuss the question of future on the socio-engineering dimension of artificial intelligence, approaching modern Confucian society's values and how humanism can be exercised in it.

In the modern intelligent information society, the role of artificial intelligence is emerging in the emergent use of information and its meaning. The point is how the engineering world of artificial intelligence relates to the world of autonomous subjects. There is an incompatible inherent problem of humanity, such as machine's humanization and human's mechanization.

Artificial intelligence discourse can be approached from either an ontological or axiological aspect of human nature's autonomous subject. According to the existing cultural theory, humans maintain and develop social relationships as they expand the scale and size of memes. The function of artificial intelligence is closely related to the role of non-genetic replicators, like this memetic method.

From the socio-engineering dimension, this is to continually acquire and evolve social and cultural aspects through integration and coordination, with the ability to self-control. This function has the persistence of autonomous order through the bonding of social structures to humans. This is in line with the sense of community in Confucian society.

In a Confucian society, humans can self-regulate themselves and form a sense of community. For example, it means based on the neutralizing mechanism of 陰 and 陽 in human relations, that is, the interaction of homogeneity and heterogeneity, avoiding conflicts and confrontations between oneself and others, oneself and the community through virtues such as 仁 and 義, 誠 and 敬, and so on. It is to move toward a sustainable horizon of communication, symbiosis, growth, and prosperity based on a bond that adjusts and integrates according to social empathy.

Thus, at the socio-engineering dimension of artificial intelligence, in the process of adjusting and integrating causal decisions and choices of free will, one can constantly create the "living force of creation" while constantly forming the "life creativity" in the continuum of Discernment and integration, office and coordination, differentiation and communication. In conclusion, I think this will be an opportunity to explore the encounter between tradition and modernity, humanities and science, philosophy and engineering, and philosophical and scientific thinking. It contains hopeful prospects and meaningful expectations for a humanism that 21st-century philosophy should bear.

본고에서는 4차 혁명시대에 인공지능의 담론에 주목하고 인공지능을 현대사회에서 어떻게 수용해야 하는가 하는 문제의식에서 출발한다. 여기에 접근하는 방법론적 발상으로서, 인공지능의 사회공학적 차원에서 현대의 유교사회의 가치관에 접근하고 그 속에서 휴머니즘이 어떻게 발휘될 수 있는가 하는 命運의 문제를 논하고자 한다.
현대의 지능정보사회에서 정보와 그 속에 담긴 의미를 창발적으로 활용하는 데에 인공지능의 역할이 대두되고 있다. 그 논점은 인공지능의 공학적 세계가 자율적 주체의 세계와 어떻게 관련되는가 하는 점이다. 여기에는 기계의 인간화와 인간의 기계화와 같은 양립할 수 없는 인간성의 내재적인 문제가 있다.
인공지능의 담론은 인간성의 자율적인 주체에 대한 존재론적 측면이나 가치론적 측면에서 접근될 수 있다. 기존의 문화이론에 따르면, 인간은 밈(meme)의 규모와 크기를 확충함에 따라 사회적 관계를 유지하고 발전한다. 인공지능의 기능은 이러한 밈의 방식과 같이 비유전적 복제자의 역할과 밀접하게 관련된다. 이를 사회공학적 차원에서 보자면 스스로 자기조절의 능력을 갖고서 통합과 조정의 방식을 통해 끊임없이 사회와 문화의 양상을 끊임없이 습득하여 진화하는 작동을 하는 것이다. 이러한 기능이 인간에게 사회적 구조의 유대적 접속을 통한 자율적 질서의 지속성을 지닌다. 이는 유교사회의 공동체의식과 맞물려있다.
유교사회에서 인간은 스스로 자기조절의 능력을 갖고서 공동체의 의식을 형성한다. 예를 들어, 인간관계에서 陰과 陽의 中和的 기제, 즉 동질성과 이질성의 상호작용을 토대로 하여 仁과 義, 誠과 敬, 등의 덕목을 통해 자신과 타인, 자신과 공동체 사이의 갈등과 대립을 지양하는 것이다. 사회적 공감력에 따라 조정하고 통합하는 유대감을 바탕으로 하여 소통, 공생, 성장 및 번영의 지속가능한 지평으로 나아가는 것이다.
그러므로 인공지능의 사회공학적 차원에서는 因果的 결정과 자유의지의 선택을 조정하고 통합하는 과정에서 분별과 통합, 직분과 조정, 분화와 소통의 연속선상에서 '생명의 창조력'을 끊임없이 구성하면서 '창조의 생명력'을 끊임없이 만들어낼 수 있다. 결론적으로 말해, 이는 전통과 현대의 만남, 인문과 과학의 만남, 철학과 공학의 만남, 철학적 사유와 과학적 사유의 만남 등을 모색하는 계기가 될 것으로 생각한다. 여기에는 21세기 철학이 짊어져야 할 휴머니즘의 命運에 대한 희망섞인 전망과 의미심장한 기대가 함께 담겨있다.

SESSION 3 : Revolutionize AI

Autonomous Vehicles, Discrimination, and Age

David Černý Karel Capek Center(Czech Republic)

While autonomous vehicles promise a whole number of benefits, introducing them into traffic may also lead to some negative consequences. One of them may be discrimination based on some physical features, for example, age. At the level of values, rejecting discrimination is based on the idea that all human beings are equal in their freedom and rights. However, it may be surprisingly difficult to come up with a correct and adequate definition of discrimination. Furthermore, it might be even more difficult correctly to apply this definition within the sphere of autonomous mobility. In my presentation, I will strive to provide a semiformal definition of discrimination. After that, I will thoroughly consider and evaluate the role of age in our decision-making processes. These reflections, coupled with the definition of discrimination, will allow me to answer the question of whether distinguishing based on age is necessarily and in all situations, including life-and-death situations, an instance of discriminating.

SESSION 3 : Revolutionize AI

Moral control over AI, What it is and why it matters.

Filippo Santoni de Sio Delft University of Technology(Netherlands)

AI may create different gaps in control and human responsibility. These may happen in the form of: shifts of power, loss of meaning, accountability gaps and responsibility gaps. In order to address these gaps a new concept of “meaningful human control” is needed, which recognises control as grounded in “responsiveness to human reasons.

SESSION 3 : Revolutionize AI

The socio-technical design of trustworthy AI

Jeffrey Chan Kok Hui Singapore University of Technology and Design(Singapore)

Increasingly, predictive algorithms are being deployed in high-stakes decision-making with significant social consequences. A recent controversy in the UK, where students were unfairly graded by algorithms after examinations were canceled because of the pandemic, illustrates the risks of relying on high-stakes algorithmic decision-making. These risks have consolidated the perception that algorithms should be seen as untrustworthy until proven otherwise (O'Neil, 2020), and reinforced the imperative for a trustworthy AI. The European Commission's Ethics Guidelines for Trustworthy AI (2019) has been seen as one of the most systematic statement of this imperative (Floridi, 2019). These important guidelines primarily focus on defining the essential ethical principles that can guide the development of a trustworthy AI, for instance, in the development of an explainable (XAI) AI system (see Hagras, 2018), among other ethical standards. While these guidelines recognize the nature of AI as socio- technical systems (AI HLEG, 2019), they do not yet focus on the implications of this recognition. The primary implication of defining AI as socio-technical systems is to see them as systems that involve a complex interaction between humans, machines, and other environmental variables (see Bauer & Herder, 2009: 5). How these different interactions take place, under what material, spatial or environmental conditions and through which agency can therefore reinforce, or detract from, the complex attribute of trustworthiness. After all, trustworthiness is not merely a technical attribute; it is also a complex psychological and social quality perceived by experiencing and interacting with AI as a socio-technical system. For example, a self-driving car that satisfies all the guidelines for trustworthiness (AI HLEG, 2019), is still likely perceived to be untrustworthy when 'driven' very cautiously by the AI on a winding mountain road in a dense fog.

AI is increasingly experienced through interacting with everyday objects, from the self-driving car to the home assistant such as Alexa. In tandem, the share of such commonplace interactions with everyday AI objects is likely to grow rapidly. While an accountable, explainable and transparent algorithmic design is necessary for a trustworthy AI, these technical specifications are not sufficient. A knowledge gap therefore exists on how to also reinforce trustworthiness using the material and social design of AI. In this paper, I will first define the concept of trust in the context of AI. 'Trust in AI' is dissimilar from the everyday notion of trusting a person, and this distinction warrants further explication. Second, I will delineate three key attributes of a trustworthy AI, namely, competency, explainability and transparency, and then explain why they are necessary but not sufficient for a trustworthy AI—especially when AI is experienced as a socio-technical system. Finally, I discuss how HCI design and regulatory design are important, but under- developed considerations for a more trustworthy AI.

BIBILOGRAPHY

AI HLEG. (2019). *Ethics guidelines for trustworthy AI: High-Level Expert Group on Artificial Intelligence*. Brussels: European Commission.
Bauer, J.M. & Herder, P.M. (2009). Designing socio-technical systems. In A. Meijers (ed.), *Philosophy of Technology and Engineering Sciences, Volume 9*. Amsterdam: Elsevier, pp. 601-630.
Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1, 261-262. Hagras, H. (2018). Toward Human-Understandable, Explainable AI. *Computer*, 51(9), 28-36.
O'Neil, C. (2020). Mutant algorithms are coming for your education. *Bloomberg*, 8 September, 2020. Retrieved from: https://www.bloomberg.com/opinion/articles/2020-09-08/mutant-algorithms-are-coming-for-your-education

SESSION 3 : Revolutionize AI

AI and the Constitutions of the Future: How can AI revolutionize our social-moral order?

John Danaher National University of Ireland(Ireland)

Human values seem to vary across time and space. Sometimes these variations are precipitated by revolutions in technology. This is unsurprising. New technologies afford new possibilities for action and social organisation. Will similar changes result from the AI revolution? Will a new social order emerge in its wake? In this talk I want to present a model for thinking about this question. I suggest that the values embraced by future societies will depend on how they manage the relationships between three kinds of intelligence: individual intelligence; collective intelligence and artificial intelligence.

SESSION 3 : Revolutionize AI

Artificial Empathy in Socially Assistive Robotics. A Philosophical Exploration

Luisa Damiano University of Messina(Italy)

In the last decades, an increasing number of emerging areas of robotics – Cognitive Robotics, HRI, Social Robotics, and in particular, Socially Assistive Robotics – have been focusing on “Artificial Empathy”, that is, a line of research dedicated to produce robots capable of effectively interacting with humans through emotions. The interest of these scientific communities in Artificial Empathy relies on two main reasons. On the one hand, producing “emotional” or “emphatic” robots means to significantly contribute to the genuinely scientific goal of creating artificial models (i.e., hardware models) of natural cognitive processes – that is, a class of processes to which emotions belong, following the Embodied Approach to the Cognitive Sciences. On the other hand, competent emotional interaction appears as a crucial aspect of robots destined to interact with humans, especially in assistive domains. It is considered one of the main components of a believable artificial “social presence”, and a “social skill” that is essential for embodied artificial agents in order to engage humans in comfortable and potentially long-lasting interactions. This talk will focus on the nascent field of Artificial Empathy, with a focus on Socially Assistive Robotics, with a twofold intent: defining the paradigmatic views of emotions underlying current research engaged in creating “empathic robots”; discussing the implications of these paradigms for the future of Artificial Empathy, in particularly with regard to the social sustainability of the diffusion of empathic robots.

SESSION 3 : Revolutionize AI

Love all, trust a few–on human relationship with AI

Saša Horvat Rijeka University(Croatia)

The phenomenon of trust underlies human relationships. People trust each other to a certain extent in everyday tasks and joint adventures. Only when our relationship with other human beings is of more importance do we begin to formulate questions such as: how much should we trust the other person, in what way and on what ground is this trust based? Human caution regarding trust was nicely expressed by William Shakespeare when he wrote: “Love all, trust a few, do wrong to none” (All's Well That Ends Well).

When is comes to artificial intelligence (AI), there is an ongoing debate which questions the possibilities of building trust between human beings and AI. This has become an interesting topic in the EU, especially since the European Commission’s High-level Expert Group on AI in 2019 posited that under certain conditions we should trust in AI and that we need to develop a trustworthy AI.

AI has become part of human everydayness and we interact with it on daily basis. On mobiles we have popular apps that count our movements and suggest how to stay healthy; algorithms suggesting news for us; on the roads we encounter smart cars driven by AI; in hospitals doctors consult in (in)direct way with AI, as well as social workers when making a decision about someone’s future and the probability of whether that person will repeat violent behaviour. Although we are becoming used to certain levels of interaction with AI, one thing is to trust mobile application on how much walking is good for you and another is to trust and decide about someone’s future and freedom with the help of probability statistics obtained from large data samples analysed by algorithms. These all are different kind of ways we interact with AI that has different kind of consequences and responsibilities. Further on, things can get even more complicated in our relationship with AI. What about complex systems where software is based on deep learning in reaching its decision or executing an action? In this case, we could be facing a lack of explainability of machine’s learning process that had led to a certain decision/action. Can we trust a “black box”? For example, this could be of crucial importance for doctors when they are making a diagnosis with the help of AI.

Therefore, we can argument that talking about trust and AI is not a black-and-white situation — either we trust AI or not. It is a complex relationship, that requires further investigation. There are other very important factors and questions in play here, for instance: is it justifiable to talk about relations with AI using concept of “trust”, since this is primarily a phenomenon of human relationships, or, in another words, can we anthropomorphise AI just for the sake of building a long-term social bond between humans and AI?

The aim of this lecture is to provide philosophical analysis of the fundamental factors of the development of trust between people and on this basis assess what it is that we really mean and want when we talk about trust in AI? This will provide us with a better insight into the complex issues of the phenomenon of trust and offer certain perspectives on future developments of the relationship between people and AI.

SESSION 4 : Human and Non-Human City

Forget being (only) smart and start being (bio)ethical! Weighing pros and cons of technology use in urban settings

Iva Rinčić · Amir Muzur Rijeka University(Croatia)

The early history of the smart city concept started back in the late 1960s when the Community Analysis Bureau in Los Angeles commerce program of “a comprehensive analysis of the entire city” produced a final document titled „*The State of the City: A Cluster Analysis of Los Angeles*,“ 1974; Vallianatos, 2015). The idea behind it was to provide enough raw data for future city strategies of various cluster analyses within urban settings. In the following decades, use and application of modern technologies (specially artifical intelligence) was extremely expanded, imposing „smart“ as a desirable aim for politicians and a bussiness oportunity for (private) investors. Hence, the term „smart“ itself remained misted and doubtfull, non-critically magnifying data-driven decisions and a technological value-free approach to city legacy. The aim of this presentation it to rethink our common understanding of the relationship between city and technology (using the example of smart cities), to critically evaluate the term „smart,“ and to offer arguments for more social, cultural, and (bio)ethical priorites for urban living.

SESSION 4 : Human and Non-Human City

Zombification of humanity - a pandemic related phenomenon

Srećko Gajović University of Zagreb(Croatia)

The term “zombification”, as established by a series of modern works of fiction, can be used to describe the transformation of dead into undead, but as well to address the contagious and irreversible change of human to un-human or zombie. The zombification as infection spreads fast and overall in a pandemic way. As the consequence of zombification, the humanity is to be divided in two groups, un-infected humans and infected un-humans, the humans being in the constant danger to be dragged in un-humanity by un-humans 1).

In the first phase of COVID-19 pandemics it was unclear how long a person would stay infected, and if those infected would stay carriers for extended periods of time representing a prolonged epidemiological hazard for their surroundings. Later on, the infection was shown to be limited in time, however some fears of reappearance of virus, or preserving the virus in the human body on the small scale (e.g. in the nervous system) remained. Subsequently, the first assumption of COVID-19 caused zombification by the permanently infected and subsequently dangerous individuals turned to be false.

However, although not permanently, the infected persons, in particular asymptomatic carriers, represent a danger for not-infected peers, supporting the aspect of zombification 2). There is a contrast of dangerous vs. in danger entities and a constant transformation of “in danger” pool into “dangerous” pool by the dangerous “agents”. The ethical deviation of the zombification process, where the citizens acquire and carry label of “infected”, hence after zombification are in need to be isolated and controlled are evident. Nevertheless, the societal mechanisms to control the pandemics by aggressive epidemiological measures are created, installed, applied and to be kept for future “challenges”.

The contrast of zombies vs. humans can be relevant to less obvious gradients of zombification where the idealized humans characterized by a set of flawless features turn to be in danger to be corrupted by not-so-perfect fellow citizens 3). These “imperfections” can be “diagnosed” by specific characteristics, and subsequently be a target for “exclusion” and denying the carriers some level of societal rights.
Artificial intelligence (AI) although having a human origin, is clearly labeled as non-human (i.e. artificial) and currently is in the status to be isolated and controlled in the technological niches. The attribute of danger is already assigned to AI. However, as AI has no infective features to humans, and it is not self-propelling, it classifies as “dead”. A speculation of transformation of “dead” AI into an “undead” AI with zombie-like features could be imagined. Although the current humanity, shattered by pandemic and corona-crisis, appreciates the isolation and exclusion strategies, the digital society already involves technologically created non-human entities but with human-like features. The humanization process applied to technology meets in the middle its reversal being the zombification process of humanity. Moreover, the blurred approach to zombification could be imagined as a societal mechanism directed versus different representations of the “undead”.

BIBILOGRAPHY

1) Nugent C, Berdine G, Nugent K. The undead in culture and science. Proc (Bayl Univ Med Cent). 2018;31(2):244-249. *doi:10.1080/08998280.2018.1441216*
2) Hirshbein L. The Zombie Apocalypse and COVID-19 [published online ahead of print, 2020 Jun 17]. Acad Psychiatry. 2020;1-2. *doi:10.1007/s40596-020-01260-7*
3) O'Brien M. You are my death: the shattered temporalities of zombie time. Wellcome Open Res. 2020;5:135. Published 2020 Jun 10. *doi:10.12688/wellcomeopenres.15966.1*

SESSION 4 : Human and Non-Human City

Walking in the City in the Age of Artificial Intelligence

Taesik Kim Masaryk University(Czech Republic)

This presentation aims to raise questions on urban practices that are possibly affected by AI-integrated urban development. As Michel de Certeau (1984) claimed that “the act of walking is to the urban system what the speech act is to language or to the statements uttered,” (p. 97) walking is the primary means to activate everyday communications in urban life. By walking, people experience urban space in many different ways, which contributes to the social production of space (Lefebvre, 1991). However, our lived experiences in urban space have dramatically changed as various communication technologies have been integrated into our urban practices. Media platforms nestled in mobile technologies are now the loci making sense of place. Algorithm-integrated technologies have vastly altered the lived experiences of urban users. Thus, this presentation brings out the first question; how would the AI-integrated urban development change the process of the social production of urban space? Urban landscapes as discursive and symbolic images often reflect various dominant and marginal values in given societies. Considering the media discourses that are heavily influenced by dominant values, this presentation raises the second question: how would the AI-integrated media technologies alter the dominant urban images that conceal multiple layers placed in urban landscapes? The relationship between technological advancements and urban practices is well reflected by the discourse over the smart city, which is limited within the functional efficiency of a city. In this discourse, the social space of lived experience is often structured by technology-integrated urban planning. The last and the most fundamental question of this presentation is grounded in this concern: how would we experience our everyday space in the age of AI?

BIBILOGRAPHY

De Certeau, M. (1984). *The Practice of Everyday Life.* Berkeley: Uuniversity of California Press.
Lefebvre, H. (1991). *The production of space.* Blackwell: Oxford.

SESSION 4 : Human and Non-Human City

A.I. and Mental-Health: A Paradigm of Human Mechanisation

Aditya Nayak Jawaharlal Nehru University(India)

Since the industrial revolution, technological research catering to the needs of the industry has attempted to reproduce human effects through machines. The machines were imagined to become humans. In the contemporary age of Artificial Intelligence, machines seek to reproduce not just the physiological human functions, but also attempt a cognitive reproduction. Unlike the existing position on automation that machines replace or supplement human labour, the paper proceeds with the understanding that the machines are strategically placed in a way that enables further extraction of labour power from humans (Nayak, 2020).

This paper argues that alongside the process of humanisation of machines, there has been a parallel process towards a mechanisation of humans. This constitutes a complementary dual- process. Because of neuro-plasticity, our neural networks are destroyed and new neural pathways are formed in the process of learning/training (Malabou, 2009). This is much like re-engineering of machinic processes.

There are several ways through which our neural pathways are structured, and the process of structuring can be controlled to precision. This research looks in particular at the contemporary mental-health discourse as a formative element of the process which seeks to re-structructure or train humans to imagine themselves in mechanised ways. The argument is that the dual process seeks a unity of human and the machine. There exists a synchronised dialogue between Artificial Intelligence implementation around us, and the mental health discourse.

The mental-health paradigm establishes a “pathology” (Canguilhem, 1989) and also institutes a system of treatment. The argument is not so important on its own, but when seen as a process that is complementary to Artificial Intelligence implementation, the argument becomes much more significant. It opens a possibility of understanding the emerging emphasis of mental-health as a controlled phenomenon that is re-engineering our neural pathways in order to prepare us for the future where A.I. enables behavioural profiling, ‘automatic-spying’ (Stiegler, 2016), algorithmic-governmentality (Rouvroy & Berns, 2013). Mental-health then becomes an instructive influence of the “desiring-machine” (Deleuze & Guattari, 1975) on human perception and reality.

The accelerated speed of existence (Virilio, 1977) and a continuous bombardment of technologically-mediated distractions leave us with a neurotic condition. While it is commonly believed that mental-health helps us in dealing with this neurosis and achieving normality, this paper argues on the contrary that mental-health keeps enforcing neurosis as a way of maintaining normality. We are expected to surrender our cognitive faculties and submit to the regime of machines, eventually becoming machines.

BIBILOGRAPHY

Nayak, A. (2020). Artificial Intelligence, automation, and changing forms of work. *Journal of AI Humanities, 5*, 57-76.

Malabou, C. (2009). *Ontology of the Accident: An Essay on Destructive Plasticity.* Cambridge, USA: Polity Press.

Canguilhem, G. (1989). *The normal and the pathological.* New York: Zone Books.

Stiegler, B. (2016). *Automatic Society: The Future of Work* (Vol. 1). Cambridge: Polity Press.

Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation. Disparateness as a precondition for individuation through relationships? *Réseaux, 177*(1).

Deleuze, G., & Guattari, F. (1975). *Anti-Oedipus.* University of Minnesota press. Virilio, P. (1977). *Speed and Politics.* France: Semiotext(e).

SESSION 5 : The World Proceeded by Artificial Intelligence

Intelligence IO

Dobrivoje Lale Eric Center for the Promotion of Science(Serbia)

A sad 2020 showed so many sums of many of our fears. It also emphasised multiplied shortcomings, misconceptions and prejudices, even corruptions and superficialities at highest levels. It also undermined our belief in Knowledge, as a humankind and as a civilisation capable of astonishing achievements. Our technology-powered and knowledge-based societies literally collapsed, thanks to something with just a couple of features. We instantly forgot past − days, years, even decades of progress − getting (b)locked by something unseen. Our systems of belief have been hacked by a clearly alternative way of thinking and better applied intelligence. That intelligence is binary, and it has only positive and negative solutions. Not to exist isn't really an option and we should try to observe ourselves through that principle.

The ultimate goal of artificial intelligence (AI) would be to create all-encompassing means capable of solving any issue and coping with the greatest of all challenges that we're facing. It was somehow foreseen that AI will represent human+ intelligence. But what if for our world, or for all the worlds that we're capable of imagining there is a more adequate protagonist, or − better to say − more compatible option? Less destroyer, more upgrader.

Our **AI Lab** has been growing yet we were forced to halt its progress and reconsider ourselves. In a globalised world, even a tiny, neglected subject from a remote market booth may become a prophet of new paradigm, or a media star. Therefore, we need to be humans more and to deploy all our knowledge and imagination. The AI Lab is fully open in 2020 as we can't offer anything determining and closing. No protocols, just collaboration and solidarity.

Fifth edition of the **art+science** programme of the Center for the Promotion of Science took place in Trieste (Italy), Belgrade (Serbia), and Linz (Austria), as well as – predominantly – in numerous virtual environments. In August and September of 2020, it brought to the audience inspiring and critical artworks created by local artists and transdisciplinary project teams engaged in particular topics within the vast field of artificial intelligence. The main event of the programme, the exhibition **Intelligence IO**, was conceived and conceptualised with the idea to tackle, analyse and discuss a very unique moment of our existence. It offered alternative perspectives and analyses of the relations between AI and different aspects of our (global) communities, including environmental and climate threats, spiritual heritage, safety and controlling mechanisms, labour framework etc.

CPN has been continuously working on creating a sustainable network which would gather scientists and artists, encourage their collaboration, and support development of their joint projects. The *a+s+cpn selection* initiates and supports interdisciplinary collaboration and the creation of new connections, moving beyond, or erasing, boundaries between actors on both sides. The Center provides resources, as well as, professional and logistic support during the evolution and development of projects. The focal point of the each art+science edition belongs to the award-winning work of the national *art+science (AI Lab) selection*. In 2020, through the project *Digital prayer*, young artist Kristina Tica established a connection between the canonical structure of the Orthodox icon and the image artificially generated by means of a computer program. The work is created and executed through the collaboration with the Mathematical Institute of the Serbian Academy of Sciences and Arts.

Center exhibits as well projects of local artists, or artists originating from Serbia, whose work has achieved significant recognition and success on global scale through the *national+ selection*. The art+science events usually offer insights into current trends and major achievements at the global scene, unlike in 2020 due to global pandemic and related restrictions. However, in 2017, one of international selections brought to Belgrade the Korean duo *Shinseungback Kimyonghun* who presented their pieces *Flower* and *Cat or Human* at the exhibition in the Cultural Centre of Belgrade. The program was supported by the Korean Embassy in Belgrade.

AI needs a critical societal analysis and valorisation, and its values and meaning should be integrated into formal educational processes, however informal educative formats targeting all ages should be stimulated as well. It could even become a driving force of the STE(A)M educational framework. Having said that and taking into account what was realistic to achieve in 2020, the exhibition Intelligence IO is more of a sequence in time than a firm, compact event. By means of limitless artistic creativity, firm scientific foundation and advanced technologies embodied through AI systems, engaged artists, scientists, educators and researchers with diverse backgrounds were unified around the idea to tackle, open up and discuss a very unique moment of our existence. The results were presented in numerous ways and − on the safe side − virtually. This, however, leads to an existential question − how to treat digital content which obtained its physical representation only for being re-transferred to a virtual realm? In a loop that may announce new modes of (co)existence, we shall consult the classic − (being) replicant or not? Exist or not? IO

* The art+science programme is a constitutive part of the Center's engagement within the European ARTificial Intelligence Lab − AI Lab project, financed through the EU Creative Europe programme. The project consortium involves thirteen partners across Europe, led be the Ars Electronica from Linz, Austria.

SESSION 5 : The World Proceeded by Artificial Intelligence

Artificial Intelligence, Personal Decisions, Consent, and the Confucian Idea of Oneness

Pak-Hang Wong University of Hamburg(Germany)

The pervasiveness of AI systems has brought forth a new background condition I call “the interconnectedness condition”, where every individual is tightly and seamlessly interconnected. In this talk, I show that personal decision-making and consent have acquired new moral significance due to the changing moral character of these acts in the interconnectedness condition. In particular, I argue personal decision-making and consent are transformed from self-regarding acts to other-regarding acts, and that the transformation introduces a new moral responsibility for individuals qua users of AI systems to account for others’ values and interests in making personal decisions and giving consent. The new responsibility, however, can be difficult for proponents of Western ethics and political philosophy to understand and accept. Accordingly, I turn to the Confucian idea of oneness to make sense of the new responsibility in the interconnectedness condition.

SESSION 5 : The World Proceeded by Artificial Intelligence

An argument against the unlimited applicability of Artificial Intelligence in classroom settings

Thomas Sukopp University of Siegen(Germany)

In my talk I will basically argue against a certain claim that makes AI a kind of universal supertool for educating young students. Obviously, AI is relevant in many respects in classroom settings. I do not promote any technophobic view on the application of modern media; of course, robots and computer programs might support young learner's competencies and even promote knowledge gaining. Nevertheless, a lot of research overestimates rather technical aspects and consequences of its use. Is AI really able to initiate more than processes related to gathering and structuring information? Is it an adequate tool and paradigm to initiate higher level education?

I will proceed as follows:
(1) Definition of AI
(2) Why AI is important and even indispensable in classroom settings in general
(3) An argument from philosophy of education/ anthropology/ethics that shows basic limitations of AI
(4) Conclusion

Ad (1): Definitions of AI are - of course - highly disputable. From my point of view such a definition should differentiate between strong an weak forms of AI, where strong AI implies systems (e.g. machines) with minds whereas weaker forms of AI only try to simulate intelligent behavior, i.e. typically problem solving. With respect to section 3 we definitely need strong AI for educational purposes.
Ad (2): I only list a few functions of AI in classroom settings without further explication: a) AI might improve communication skills; b) AI helps to gather information; c) AI helps to design curricula; d) AI helps teachers to grade their students/pupils.

Ad (2): I only list a few functions of AI in classroom settings without further explication: a) AI might improve communication skills; b) AI helps to gather information; c) AI helps to design curricula; d) AI helps teachers to grade their students/pupils.

Ad (3): The following premises are indispensable for my argument: a) There is a deep gap between AI systems that simulate a human capability such as empathy, teacher's authority, openness and receptiveness, and e.g. real empathy. From a phenomenological point of view, young students feel these differences. b) Some dimensions of human acting (as a teacher) can only be conveyed by real human beings. (To avoid a petitio-conjecture I will explain why (b) is a proper assumption.) One example: At least from a phenomenological point of view there is a clear difference if a person educates pupils/students with his "pedagogical eros" instead of robots. c) AI cannot cover these dimensions. In conclusion, AI cannot educate pupils/students in many respects in the full meaning of upbringing and education as "Bildung". To strengthen my argument I will focus on basic aims and goals of education. "The educated person relates the legitimate demands of our world to one another as well as the purposes being valid and established." (Ladenthin 2007, 96). Moreover, irreducible to economic, social or political factors, an individual creates in this whole process (including the process of education as "Bildung" and upbringing as well as the process of learning) meaning. AI systems are far away from being able to achieve these goals.

Ad (4): We should keep in mind that the use of AI has fundamental limitations. Therefore, it would be short-minded to sacrifice dimensions of learning and education simply because of some hegemonic claims of AI-admirers.

SESSION 5 : The World Proceeded by Artificial Intelligence

Does Knowledge Gap Theory Explain the Difference?: An Empirical Study of Artificial Intelligence and Its Impact on Human's Perception

Geonwoo Park Konkuk University(Korea)

This study aims to empirically investigate people's perceptions of artificial intelligence on its impact. In the era of the 4th Industrial Revolution, the change in many aspects is discussed a lot. The pieces of evidence from this change are easy to find. For example, the automated robot with artificial intelligence in the manufacturing sector is a common phenomenon these days. With this observation, the researchers try to find what is the benefit and harm of this change. However, it is still underdeveloped that the general perception of the people on this knowledge. To fill the gap, this study mainly handles how people think about artificial intelligence and its use and result. This study uses survey data for the analysis. To find an interesting point, it focuses on the knowledge gap theory about artificial intelligence, automation, and robot of the people in South Korea. The contribution of results can be used for understanding human's perception of this dramatic change and how people respond.

SESSION 5 : The World Proceeded by Artificial Intelligence

AI and the Two Cultures

Heeryon Cho Humanities Research Institute(Korea)

We see majority versus majority opposition escalating in many areas of our society, and recommendation services could be the major force driving this phenomenon. Multidisciplinary efforts to implementing a data-driven technology can often benefit the society, so collaboration between arts and humanities and science and engineering is vital. I call for the harmonization of “The Two Cultures” in the creation of thoughtful technology.

요즘 우리 사회에서 두 집단이 강하게 대립하는 모습을 자주 보게 되는데, 이러한 현상은 추천 서비스의 영향일 수도 있다. 데이터 기반 기술이 사회를 이롭게 하려면 기술 도입 시 다양한 분야의 전문가가 참여하는 것이 중요하며, 과학과 인문학의 융합이 필수적이다. 사려 깊은 기술을 창출하기 위해 C. P. 스노우가 제창했던 ‘두 문화’를 되짚어 본다.

SESSION 5 : The World Proceeded by Artificial Intelligence

Development and Application of <Tokki-jeon Chatbot> for Education Using Open Source Chatbot

Myeonghyeon Lee · Yujin Kang Chung-ang University(Korea)

Chatbots and classical literature are apt to think of as disparate. However, if we want to apply the computing thinking that has characteristics as problem-disassemble, pattern recognition/data representation, generalization/ abstraction, and algorithms to the educ ation of classical literature, the possibility of classical education using chatbots is very high.

Currently, engineering that leads to the development of chatbots has low interest and un derstanding of classical literature education, and many of those who major in classical lit erature are ignorant of algorithms and program development. Under these circumstance s, for the efficiency of access to chatbots after research and development, this study will develop chatbots by using open chatbot sources such as Danbi and Aquinator as puzzle-matching games.

If open-source of a chatbot is used, a chatbot can be made without having to develop alg orithms and programs directly. However, through the process of designing and impleme nting chatbots, the characteristics of algorithms can be understood, and computing thin king can actually be applied to the understanding of classical literature through concept definitions, processes, sequences, and arrangements.

챗봇과 고전문학은 서로 이질적인 것으로 생각하기 쉽다. 그러나 문제분해, 패턴 인식/자료 표현, 일반화/추상화, 알고리즘을 특성으로 하는 컴퓨팅 사고를 고전문학 교육에 적용하고자 한다면 챗봇을 활용한 고전문학 교육의 가능성은 매우 높다고 할 수 있다.
현재 챗봇 개발을 주도하는 공학에서는 고전문학 교육에 대한 관심과 이해도가 낮고, 고전문학를 전공자들은 알고리즘, 프로그램 개발 등에 문외한인 경우가 다수이다. 이러한 상황에서 연구와 개발 이후 챗봇 접근의 효율성을 위하여 단비, 아키네이터 등 공개된 챗봇 소스를 활용하여 퍼즐 맞추기 게임 방식으로 챗봇을 개발하고자 한다.
챗봇 오픈 소스를 활용할 경우, 알고리즘과 프로그램을 직접 개발하지 않아도 챗봇을 만들 수 있다. 그렇지만 챗봇 설계와 구현 과정을 통해 알고리즘의 특성을 이해하고, 개념 정의와 프로세스, 순서 및 배열 등을 통해 컴퓨팅 사고를 고전문학 이해에 실제로 적용할 수 있다.

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

When Artificial Intelligence Meets Neuroscience: Social, Ethical, and Legal Implications of AI-powered Neurotechnology

Anita Sue Jwa Stanford University(U.S.A)

Recent advances in artificial intelligence (AI) have transformed a wide spectrum of scientific fields. Neuroscience is one of the disciplines in which AI algorithms are being increasingly utilized and that can greatly benefit from the power of AI algorithms (Savage, 2019). Neuroscience, the scientific study of the human brain and the nervous system, has been rapidly developing during the last few decades with the aid of various innovative methods to monitor and record neuronal activation, such as functional magnetic resonance imaging(fMRI). AI technology is expected to further facilitate new discoveries in the structure and function of the human brain, for instance, by allowing researchers to process the massive data collected through these tools and parse neuronal activation from the complex data (Hong Jing, 2020; Fan, 2019).

Brain-computer interface (BCI) is an excellent example of AI-powered neurotechnology. BCI records neural signaling from the brain and translates it to control external devices. Combining AI with BCI has remarkably improved the speed and precision of decoding neural signals, and BCI has been gaining momentum as a tool to assist – and potentially to augment – sensory, motor, and cognitive function (Zhang et al., 2020). A number of neurotechnology start-ups, such as Neuralink founded by Elon Musk, are developing BCI devices, which have electrodes implanted into the brain, to read brain activity and/or modulate the activity (Lopatto, 2019). Scientists at Facebook also are currently investigating a BCI technology to translate brain activity into text so that people can type merely by thinking (Constine, 2017). BCI research in the academic environment is more geared toward neuroprosthetics, which aims to restore impaired sensory or motor functions, for example, by enabling a paralyzed patient to move a robotic arm (Vilela & Hochberg, 2020).

AI has also accelerated the application of more conventional tools in neuroscience research, such as fMRI or electroencephalogram (EEG), in decoding brain activity. Some studies were able to reconstruct visual images that were displayed to subjects through AI-powered analysis of fMRI or EEG data collected while the images were shown to the subjects (Huth et al., 2016; Nemrodov, Niemeier, Patel, & Nestor, 2018). The accuracy of decoding mental state (e.g., happiness or sadness) using these technologies has also been substantially improved by AI techniques (Kim, Kim, Oh, & Kim, 2013). Commercial EEG devices are already on the market and advertised to the general public for various wellness purposes – for example, monitoring a user's attention level and providing feedback to improve it (Wexler & Thibault, 2019).

For all its promise, however, the use of AI to decode brain activities would also give rise to important ethical and legal concerns. First and foremost, it is crucial to note that AI-powered brain decoding technology is still at an exploratory stage. Premature application of this technology could lead to detrimental outcomes. Despite some promising results reported in previous studies, existing mind-reading technologies have not yet proven to be reliable, and thus, more extensive scientific validation is needed to be used in real-life settings (Giattino, Kwong, Rafetto, & Farahany, 2019). Recent attempts to apply it to test the memory of a suspect in criminal cases or to screen applicants in hiring processes are extremely alarming as they are apparent misuses of unproven technology(Giattino, Kwong, Rafetto, & Farahany, 2019). Moreover, even if this technology becomes mature enough to generate highly reliable and accurate results, decoding neural activities could significantly infringe on our most intimate and personal area – our mind.

Given the unprecedented pace of technological advancement, we should start contemplating on how to protect privacy rights and freedom of thought from potential unwanted intrusion by AI-powered neurotechnology (Farahany, 2012; Farahany 2019). The fact that this technology can be applied not only to read neuronal activity but also to potentially modulate it poses another threat to our autonomy, especially regarding the right to change our minds or to make decisions to govern ourselves (Drew, 2019). A BCI device that reads a user's brain activity through an AI algorithm and writes back to the brain based on this interpretation could unintentionally or even intentionally intervene in the user's mental state and ultimately alter behavior, which is the outward manifestation of the mental state. This would raise interesting and important questions concerning the user's agency and accountability regarding his or her actions (Drew, 2019). Considering these potential risks, it is critical to accomplish ethical integration in advancing neuroscience research with AI techniques. Policy makers and experts across various disciplines including neuroscience, computer science, bioethics, and law should pay close attention to these risks and develop sound policy to promote responsible use of AI-powered neurotechnology.

BIBILOGRAPHY

Constine, J. (2017, April 19). Facebook is building brain-computer interfaces for typing and skin-hearing. https://techcrunch.com/2017/04/19/facebook-brain-interface/

Fan, S. (2019, August 8). Three invaluable ways AI and neuroscience are driving each other forward. *SingualrityHub*. https://singularityhub.com/2019/08/08/three-invaluable-ways-ai-and-neuroscience-are-driving-each-other-forward/

Farahany, N. A. (2012). Incriminating Thoughts. *Stanford Law Review, 64*(2), 351-408. Retrieved from http://www.jstor.org/stable/41511094

Giattino, C. M., Kwong, L., Rafetto, C., & Farahany, N. A. (2019). The Seductive Allure of Artificial Intelligence-Powered Neurotechnology. Paper presented at the Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA. https://doi.org/10.1145/3306618.3314269

Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A. T., & Gallant, J. L. (2016). Decoding the Semantic Content of Natural Movies from Human Brain Activity. Front Syst Neurosci, 10, 81. doi:10.3389/fnsys.2016.00081

Jing, H. (2020, March 3). Fascinating relationship between AI and Neuroscience, Medium. https://towardsdatascience.com/the-fascinating-relationship-between-ai-and-neuroscience-89189218bb05

Kim, M. K., Kim, M., Oh, E., & Kim, S. P. (2013). A review on the computational methods for emotional state estimation from the human EEG. Comput Math Methods Med, 2013, 573734. doi:10.1155/2013/573734

Lopatto, E. (2019, July 16). Elon Musk unveils Neuralink's plans for brain-reading 'threads' and a robot to insert them. The Verge. https://www.theverge.com/2019/7/16/20697123/elon-musk-neuralink-brain-reading-thread-robot

Nemrodov, D., Niemeier, M., Patel, A., & Nestor, A. (2018). The Neural Dynamics of Facial Identity Processing: Insights from EEG-Based Pattern Analysis and Image Reconstruction. eNeuro, 5(1). doi:10.1523/ENEURO.0358-17.2018

Savage, N. (2019). How AI and neuroscience drive each other forwards. Nature, 571(7766), S15-S17. doi:10.1038/d41586-019-02212-4

Vilela, M., & Hochberg, L. R. (2020). Applications of brain-computer interfaces to the control of robotic and prosthetic arms. Handb Clin Neurol, 168, 87-99. doi:10.1016/B978-0-444-63934-9.00008-1

Wexler, A., & Thibault, R. (2019). Mind-Reading or Misleading? Assessing Direct-to Consumer Electroencephalography (EEG) Devices Marketed for Wellness and Their Ethical and Regulatory Implications. Journal of Cognitive Enhancement, 3(1), 131-137. doi:10.1007/s41465-018-0091-2

Zhang, X., Ma, Z., Zheng, H., Li, T., Chen, K., Wang, X., . . . Lin, H. (2020). The combination of brain-computer interfaces and artificial intelligence: applications and challenges. Ann Transl Med, 8(11), 712. doi:10.21037/atm.2019.11.109

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

Gender, bias and knowledge politics: towards new materialist reconfiguration of AI systems

Corinna Bath Braunschweig University(Germany)

Artificial intelligence, together with large amounts of data (Big Data), promises to enable previously unfeasible calculations, predictions and personalization. However, more and more examples show that these methods can reproduce and intensify existing social inequalities and stereotypes, particularly the structural-symbolic gender order. In my talk I will discuss case studies that illustrate such undesirable effects and analyze the underlying knowledge politics. Based on feminist science and technology studies I will critically reflect assumptions such as the objectivity of algorithms or the idea that data simply represent reality. Another source of criticism is the computer science community itself that struggles with questions of how to build fair and socially just algorithms (for instance the ACM Conference series on Fairness, Accountability and Transparency). Though these attempts often lack advanced inter- and transdisciplinary translations. By drawing on new materialist approaches, the aim of my talk is to unfold a theoretical understanding and practical engagement for a “response-able” (Haraway 2008, Barad 2012, 2010) socio-technical reconfiguration of AI systems: designers, users, data and algorithms.

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

A Long Way to Go for Data Protection and Privacy Regulations

Hortense Gallois McGill University, Center for Genomics and Policy(Canada)

AI has led to significant improvements in healthcare, especially in areas such as precision medicine, diagnosis tools, psychological support and help for the elderly. 1) AI technologies yet rely on accessing large amounts of information, including sensitive health-related data extracted from medical files or research participants' results. 2) Procuring a convenient and safe access to health-related data has thus become a central issue, 3) yet individuals increasingly show substantial concerns about sharing their data in the medical and clinical context. 4) In the United States, studies have shown that individuals' willingness to participate in research involving their genetic data is affected by their concerns about their ability to protect their privacy in such context. 5) Paradoxically, this lack of trust is counterbalanced by a growing popularity of direct-to-consumer genetic testing and health monitoring devices. Adequately assuring the right to privacy of citizens, especially when they share their most personal information with private companies, while facilitating access to personal data for research is probably one of the biggest challenges policy makers have to face in any country wishing to benefit from the many opportunities associated with the implementation of AI technologies in healthcare. Through the adoption of its new General Data Protection Regulation (GDPR), 6) the EU was the first in attempting to regulate AI through data protection legislation. This regulation paved the way for meaningful reforms in privacy legislation in the US and Canada.

The extra-territorial reach of the new regulation puts pressure on Canada and the US to reform their own privacy legislation. Indeed, both systems likely fall short of some of the new requirements set by the EU regulation. If these laws are found to provide insufficient protection, the result could be a decrease in data flow from the EU to North America, due to the need to proceed via the adoption of additional contractual clauses. 7) Such decrease would not only negatively affect research and development of AI technologies in both countries but would interfere with any attempt at cooperation in the field. 8) In this context, it appears all the more pressing to consider appropriate measures to consolidate privacy protection and promote stakeholders' trust. We will attempt to identify key avenues for policy reform, in the United States and Canada, which we contend could be inspired by the GDPR while taking into considerations its potential limitations.

BIBILOGRAPHY

1) See Andreas Holzinger, Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning, 15 The IEEE intelligent informatics bulletin 6–14 (2014).
2) See Riccardo Miotto et al., Deep learning for healthcare: review, opportunities and challenges, 19 Brief Bioinform 1236–1246 (2018).
3) See Willem Sundblad, Data Is The Foundation For Artificial Intelligence And Machine Learning Forbes (2018), https://www.forbes.com/sites/willemsundbladeurope/2018/10/18/data-is-the-foundation-for-artificial-intelligence-and-machine-learning/ (last visited Jan 23, 2019).
4) See Brian Scogland, Artificial Intelligence in Medicine: Hope or Hype? MDDI Online (2018), https://www.mddionline.com/artificial-intelligence-medicine-hope-or-hype (last visited Jan 9, 2019).

5) See Ellen W. Clayton et al., *A systematic literature review of individuals' perspectives on privacy and genetic information in the United States*, 13 PLoS ONE, 2 (2018).
6) Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). [2016] OJ L119/1.
7) See Hayley Evans & Shannon Togawa Mercer,
Privacy Shield on Shaky Ground: What's Up With EU-U.S. Data Privacy Regulations, LAWFARE, 2018, https://www.lawfareblog.com/privacy-shield-shaky-ground-whats-eu-us-data-privacy-regulations (last visited Feb 7, 2019); Susan Krashinky Robertson, *Calls grow for Canada to modernize privacy laws amid EU changes*, THE GLOBE AND MAIL, 2017, https://www.theglobeandmail.com/report-on-business/industry-news/marketing/calls-grow-for-canada-to-modernize-privacy-laws-amid-eu-changes/article35778176/ (last visited Feb 7, 2019).
8) International collaboration in AI is based on increasingly large amounts of data, including trans-border data sets. If data exchanges are hindered, the continuation or development of transatlantic research projects in the field are thought to be impeded. See Mark Phillips, *International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR)*, 137 HUM GENET 575–582 (2018).

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

Artificial Intelligence and Crime Prevention: Friend or Foe?

James Haillot Institute of Advanced Legal Studies (IHEJ-Paris)

Artificial Intelligence (A.I) has been developed in the United States as a tool for crime-prevention since the early 2000s. At first, AI was perceived as a strongly objective tool, allowing police forces to benefit from unbiased information in crime-prevention. However, the strong racial division in the United States and the lack of training of police forces have led to a prejudicial situation for minorities. Over the years, human rights advocates have argued that the use of A.I. tools for crime prevention have created a basis for systematic discrimination. In Europe, though fairly protected by strong data-protection regulations, A.I. tools for crime-prevention have been developed and tested in France and other countries (facial recognition, 'smart'-cameras...). Strong regulations have yet prevented the massive deployment of such tools, which when they were, have been vigorously criticized by the public and data-protection watchdogs.

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

Deep learning as boundary tools: perpetrating and detecting bias. The Case of COVID-19.

Lucia Martinelli Muse-Science Museum(Italy)
Jiwon Shim Humanities Research Institue(Korea)
Sungeu Chae Hanyang University(Korea)

Adopted at the end of 1700 to refer a printing process based on rigid forms [στερεος + τύπος], the term “stereotype” was firstly introduced by journalist Walter Lippmann in early 1920s to indicate “pictures in the head,” i.e. mental reproductions of reality. Today, the term evolved meaning a “cognitive schemas or beliefs used by people to process information about or guide behavior toward individuals or social groups” (Marx & Ko 2019). Worth stressing, the act of categorization of the mental images of the reality are typically very simplified and very rigid. Whilst stereotypes refer to both positive and negative overgeneralizations or beliefs, “prejudice” (closely related to bias) usually refers to the negative aspects of the stereotype, being based on *a priori* and not experienced idea. Both terms often get together to justify unfair or biased treatment toward members of a group on the basis of various features (sex, gender identity, race and ethnicity, nationality, age, socioeconomic status, language, and so forth).
Stereotypes as well as prejudices are something very human (Roets & van Hil, 2011) and deeply embedded within social institutions and wider culture at social and unconscious levels. Since cognitive bias are connatural in our thinking, they are hard to be identified and exercise is needed to gain awareness about the limited mental patterns affecting our judgments and decisions when these are based on unconscious bias (see for instance the project IMPLICIT, https://implicit.harvard.edu/implicit/takeatest.html). Humor is typically based on stereotypes, with the risk to become disagreeable and even offensive, generally bordering on racism and sexism (Bemiller & Schneider 2010). Unconscious bias can impede objective and fair judgements risking to give rise to discrimination in important contexts, as for instance in medical treatments, decisions related to criminal justice and hiring and carrier promotions (Régner et al. 2019).

The role of media in building up a collective imaginary of reality, often based on stereotypes, have been already pointed out long ago (Lippman, 1949). Nowadays, digital media, in particular with the increase in the use of mobile phones, has become closer and closer to people's daily lives allowing producing and sharing news promptly and widely. In digital media news, visual communication is a crucial corollary to the narrative delivery by emphasizing the news, attention capturing as well as affective-emotional sphere reaching. Here, journalistic photographs, by symbolizing generalities and providing mythical frames for interpretation, are significant not only for their ability to describe events but rather for their role in evoking pre-existing interpretive schema that link the viewer's memory to familiar news categories (Griffin 2004). Therefore, images have the power to deliver Authors' bias as well as to interface with our own bias, being these either unconscious or conscious. In the case of diseases and pandemics, for instance, besides health advises, pictures can portray fears well rooted in human history, desperation and anger as well as stigmata concerning their popular names ("Spanish" and "Hong Kong" flus; "Middle East" Respiratory Syndrome-MERS; "Chinese virus", as in the recent COVID-19 pandemic) and marginalization of victims (AIDS) (Cowell 2020).
If prejudice/bias is typical of human beings, are machines, artificial intelligence, immune to it? Devices based on artificial intelligence have become an integral part of our lives, from judicial systems to search engine advice. But is it possible that the algorithms at the base of these useful applications suffer from a drawback of human beings that is prejudice? Can a machine, an evolved algorithm, help, correct and even replace human judgment? But above all, will our cognitive bias be amplified or find support, if not solution, in the use of artificial expert systems?

These questions will be discussed in this presentation. Artificial intelligence will be regarded as a boundary tools, having both the risk to perpetrate bias and the know-how to detect bias. In the first part, focus will be on algorithmic bias. Due to inherent biases in data sources or discriminative intention of the algorithm developers, algorithms and decision-making based on Big Data can possibly lead to discriminative results. Since even with suppressed sensitive attributes which possibly contribute to biases, subtle correlation in the data can negatively affect algorithms, machine learning community showed effort in designing techniques for discrimination discovery in databases as well as algorithmic prevention of bias (Hajian et al. 2016). Examples of some significant gender and ethnic bias, which could be contained in machine learning algorithms, will be discussed. Specifically focusing on gender and artificial intelligence, the relevance of adopting a gendered innovation approach (Schiebinger et al. 2016) and specific policies to reduce the bias will be proposed (see for instance UNESCO 2019).
In the second part, a recent our research aimed at detecting ethnic bias in the most acknowledged digital news media of South Korea, Italy and USA will be presented. Deep learning approach has been used to analyze images of mask-wearing people during hot phases of COVID-19 pandemic, focusing on discrimination against Asians wearing masks. Unlike in traditional machine learning, deep learning algorithms such as CNN (Convolutional Neural Network) can possibly extract features from input data directly, enabling integration of feature extraction and classifier. Also, since the neural network is in stacked structure (deep learning), multiple levels of layers are expected to discover abstract semantics of the data (Guo et al. 2017). Facemask wearing has been chosen as an emblematic case study of the socio-cultural expression of this devastating health crisis. We assume, in fact, that -beyond the health protection tool- facemask wearing proposes a series of features, among these a series of stigmata towards the disease, its presumed origin and the community/ethnicity more familiar with mask wearing. Based on the image data on the Coronavirus and mask-wearing, we identified in each country specific characteristics of the social representation of people wearing mask and what differences are seen in those images, detecting stigmata mostly affecting the USA digital news.

BIBILOGRAPHY

Bemiller M.L., Schneider R.Z. (2010) It's not just a joke. Sociological Spectrum 30: 459-479. https://doi.org/10.1080/02732171003641040

Covel A. (2020) Photos from a century of epidemics. The New York Times, March 20, 2020. URL: https://www.nytimes.com/2020/03/20/world/europe/coronavirus-aids-spanish-flu-ebola-epidemics.html

Guo, T., Dong, J., Li, H., & Gao, Y. (2017) Simple convolutional neural network on image classification. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA) 721-724. DOI: 10.1109/ICBDA.2017.8078730

Griffin M. (2004) Picturing America's 'war on terrorism' in Afghanistan and Iraq. Photographic motifs as news frames. Journalism 5(4): 381–402. DOI: 10.1177/1464884904044201

Hajian, S., Bonchi, F., & Castillo, C. (2016) Algorithmic bias: From discrimination discovery to fairness-aware data mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2125-2126. URL: https://doi.org/10.1145/2939672.2945386

Lippmann W. (1922) Public opinion. The Free Press, New York, Collier-Mac Millan Limited, London

Marx D. & Ko S.J. (2019) Stereotypes and Prejudice. Oxford Research Encyclopedia, Psychology, Oxford University. DOI: 10.1093/acrefore/9780190236557.013.307. URL: https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-307

Régner, I., Thinus-Blanc, C., Netter, Schmader T, Huguet P.(2019) Committees with implicit biases promote fewer women when they do not believe gender bias exists. Nat Hum Behav 3:1171–1179. https://doi.org/10.1038/s41562-019-0686-3

Roets A., Van Hiel A. (2011) Allport's prejudiced personality today: Need for closure as the motivated cognitive basis of prejudice. Directions in Psychological Science.

Schiebinger L., Klinge I., Sánchez de Madariaga I., Paik H.Y., Schraudner M., Stefanick M. (Eds.) (2011-2018) Gendered Innovations in Science, Health & Medicine, Engineering and Environment. URL: EU citations add: http://ec.europa.eu/research/gendered-innovations/

UNESCO (2019) First UNESCO recommendations to combat gender bias in applications using artificial intelligence. URL: https://en.unesco.org/news/first-unesco-recommendations-combat-gender-bias-applications-using-artificial-intelligence

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

Regulation of Artificial Intelligence

Mesut Serdar Çekin Turkey-Germany University Istanbul(Turkey)

In 1842 a faulty horse-drawn carriage caused an accident in the USA with considerable damage. The responsible court rejected liability of the manufacturer on the grounds that it is `an infant industry`, arguing that strict liability would negatively affect its development 1) . Almost 20 years later, a train spark causes a fire in Munich, destroying a farmer's entire harvest. The formation of sparks was unavoidable according to the state of the art at that time. Although there was no strict liability at that time and the train operator met all technical standards, the court nonetheless affirmed liability by assuming a “civilistic fault” 2). We are in a similar situation today: should the regulation of AI be liberal or should it be more protective? Which standards have to be developed?

The problem is not new though. The so-called Collingridge dilemma is actively discussed in the economic sciences in particular. The question is whether a new technology should be regulated immediately and strictly, with the result that security is bought at the price of the loss of possible advantages of this technology. The other alternative is liberal regulation with the result that although the potential of the new technology is fully exploited, this technology has crept into everyday life to such an extent that it is no longer possible to avoid possible, irreparable damages.

Does history repeat itself? Or does AI create a new situation? What is special about AI? At least for the moment it is certain that a fundamental change in social life can be expected. While muscle power was replaced by machines in the first industrial revolution, human intelligence is now to be replaced (at least in some areas) by algorithms. Almost all areas of life are directly affected: from the financial industry (fintech) to home living (smart home apps). There are many approaches worldwide that aim to regulate AI. So far, however, these approaches have been rather abstract guidelines and not concrete specifications.

The aim of this study is not to present concrete solutions, as this would above all require a multidisciplinary approach. However, possible approaches are to be presented as to how AI can be regulated in general.

The general regulatory concepts must be considered. Imperative regulation, co-regulation and self-regulation. In the classic regulatory method of imperative regulation, the state monitors and controls the activities of the citizens. Self-regulation, on the other hand, leaves regulation to the market players. Between these two concepts is what is known as co-regulation. According to this, market players are supposed to monitor themselves, but the state stands above them. Because imperative regulation presupposes control and monitoring, the controlling body (the state) must at least have the same level of knowledge as the controlled market players. The information asymmetry between the state and private companies is growing steadily, however. On the other hand, the complete relinquishment of control to the market players can lead to consequences which cannot be redressed. This paper therefore aims to investigate the question of whether the co-regulation model and a risk-based approach that are already known in the area of information security, environment law and data protection law can offer suitable methods for regulating AI.

Before the regulating body pronounces bans and thus nips possible potentials in the bud, the private actors should, in consultation with the state authorities, establish their own rules because they have the better specialist knowledge. The state is the supreme supervisory authority, while the professional associations monitor their own members. The advantage is that potential can be tapped without unnecessary interventions. On the other hand, the exchange of information between the state body and the professional associations must be very good in order to prevent information asymmetry. Another argument in favor of such an approach is that the so-called tech giants mainly come from the USA and self-regulation is very widespread there, while classic imperiative regulatory culture is less known to these companies. In order to avoid such a culture clashes between the different regulatory models, co-regulation can offer a suitable alternative.

Another approach can be the mandatory risk assessment. Before a product with AI is launched on the market, the respective developer should carry out a risk assessment. This approach is actually nothing new. Numerous models already exist in environmental protection law, in finance and data protection law. Only when the developer can reduce the risk to an acceptable level should the product be introduced. Since there is currently a lack of technical standards for AI, the risk of a misjudgment is entirely transferred to private actors, who have the better specialist knowledge. This also significantly reduces the problem of information asymmetry. The already developed ethical guidelines such as transparency, security, explainability, etc. can then be subordinated to this risk-based approach. The result is an adaptable and flexible regulatory approach that will be very important for the development of AI in particular.

BIBILOGRAPHY

1) Winterbottom v. Wright (1842) 10 M. & W. 109, 11 L.J. Ex. 415.

2) OAG München, Seufferts Archiv für die Entscheidungen der obersten deutschen Staaten 14, Nr. 208, sh. 355.

SESSION 6 : Calculating Good and Evil with Artificial Intelligence

The Future of AI and the Society of the Future: Law, Ethics, and the Governance of AI

Ugo Pagallo University of Turin(Italy)

The talk focuses on the future of AI and the society of the future, in light of the normative challenges brought about by this vibrant field of technological innovation. Special attent ion is drawn to the moral, legal, and political challenges of AI, and how policy makers an d institutions aim to tackle them. New forms of co-regulation, a robust set of legal amen dments and new moral principles, such as the principle of explainability, should help us address what's unique to the normative challenges of AI.