MI: Robotikától a filozófiáig

A jövő intelligens robotjai – avagy az emberi evolúciós fejlődés mesterséges intelligencia alapokon

AI: From Robotics to Philosophy

The Intelligent Robots of the Future – Or Human Evolutionary Development Based on AI Foundations

A Magyar Tudományos Akadémia és az Óbudai Egyetem által rendezett workshop anyagai, szerk. Molnár György és Nyíri Kristóf

Proceedings of a Hungarian Academy of Sciences / Óbuda University workshop, edited by György Molnár and Kristóf Nyíri

Budapest 2023

MI: Robotikától a filozófiáig. A jövő intelligens robotjai – avagy az emberi evolúciós fejlődés mesterséges intelligencia alapokon —— AI: From Robotics to Philosophy. The Intelligent Robots of the Future – Or Human Evolutionary Development Based on AI Foundations. Abstracts / papers prepared for a Hungarian Academy of Sciences / Óbuda University blended (online / physical) bilingual (Hungarian / English) workshop held on Nov. 17, 2023.

Az absztraktok / előadások szerzői —— Authors of the abstracts / papers: CSÉPE, Valéria (Prof. Dr., Full Member of the Hungarian Academy of Sciences [HAS], Member of the Presidency of HAS); EIGNER, György (PhD, Associate Professor, Óbuda University [ÓU], John von Neumann Faculty of Informatics, Dean); FALUS, András (Prof. Dr., Full Member of HAS); FROHOCK, Richard R. (graduate student at the University of South Florida); GALAMBOS, Péter (Deputy Director of the University Research and Innovation Center, ÓU); GIGERENZER, Gerd (Prof. Dr., Director of the Harding Center for Risk Literacy at the Max Planck Institute for Human Development, Berlin); KATZ, James E. (PhD, Dr. h.c., Feld Professor of Emerging Media; Division of Emerging Media Studies, College of Communication, Boston University); MOLNÁR, György (Prof. Dr., Dean, Kandó Kálmán Faculty of Electrical Engineering, ÓU); NYÍRI, Kristóf (PhD, Dr. h.c., Full Member of HAS; Chair, Committee for Communication and Media Theory, HAS; Research Professor at ÓU); SMITH, Barry (Professor in the University at Buffalo); TURNER, Stephen (Distinguished University Professor and Graduate Research Professor at the Department of Philosophy of University of South Florida); SZATHMÁRY, Eörs (Prof. Dr., Full Member of HAS, Member of the Presidency of HAS; Professor of Biology, Institute of Evolution, Centre for Ecological Research; Institute of Biology, Eötvös University; Parmenides Center for the Conceptual Foundations of Science). A workshopot moderálta / workshop discussion directed by György MOLNÁR.

Budapest: November 2023

Tartalom / Contents

Molnár György –
Nyíri Kristóf

Előszó

Jelen füzet a Magyar Tudományos Akadémia és az Óbudai Egyetem által a 2023-as Magyar Tudomány Ünnepe sorozat keretében rendezett Mesterséges Intelligencia workshop anyagait tartalmazza. Vegyes – online és fizikai – valamint kétnyelvű – magyar és angol – eseményre került sor, 2023. nov. 17-ikén. A részleteket és az eseményt megelőzően feltöltött anyagokat ld. a http://www.hunfi.hu/nyiri/AI/Opening_page.pdf oldalon és az oda kapcsolódó linkeken. A füzet – ld. a tartalomjegyzéket – egyenetlen: valamennyi résztvevő küldött absztraktot, de az eseményt megelőzően csak néhányan fogalmazták meg előadásukat. Ám mindösszesen a füzet kimagaslóan érdekes, új gondolatokban abszolút gazdag.

György Molnár –
Kristóf Nyíri

Preface

The present leaflet contains materials composed for a Hungarian Academy of Sciences / Óbuda University AI workshop that took place in the framework of the 2023 Hungarian Science Festival. It was a blended (online as well as physical) bilingual (Hungarian and English) event, held on November 17, 2023. For details, and for the materials as uploaded prior to the event, see http://www.hunfi.hu/nyiri/AI/Opening_page.pdf and the pages linked to it. The leaflet – compare the table of contents – is uneven, with all participants providing abstracts, but only a few sending in their papers prior to the event. But altogether the leaflet is highly interesting, and absolutely rich in ideas.

MAGYAR SZEKCIÓ

ABSZTRAKTOK

Molnár György

Humán-Gép interakciók evolúciója – MI alapú innovatív tanulástámogató rendszerek és tanulási környezetek az oktatásban

Az elmúlt néhány évtizedben a mesterséges intelligenciával kapcsolatosan jelentős előrelépéseket tettek olyan eszközök és robotok kifejlesztésében, amelyek képesek az emberekkel különböző kontextusokban interakcióba lépni. Ide sorolhatók a humanoid vagy szociális robotok, amelyeket számos környezetben használnak, az oktatástól az egészségügyön át a szórakoztatásig és a kutatásig. Ezeknek az intelligens eszközöknek a hatékonysága azonban nagymértékben attól függ, hogy képesek-e értelmes és vonzó módon interakcióba lépni az emberekkel. Az ember-gép interakció (HMR) egy interdiszciplináris kutatási terület, amelynek célja az ember és a gépek közötti kommunikáció, együttműködés és kapcsolatok mélyebb megértése. Az elmúlt években a HMR-kutatás eredményeit tekintve jelentős lépéseket tettek, amelyek alatt érthetjük az érzékelés, a megismerés, a navigáció és a kommunikációs technológiák terén elért eredményeket, amelyek elősegítették az ember-gép együttműködés hatékonyabbá tételét. Az ember–gép interakció egyik fő célja, hogy az ember és a gép hatékonyan tudjanak együttműködni a különböző feladatok elvégzése érdekében. A HMR-kutatás több területre összpontosít, például a szociális interakciókra, az együttműködésre, a tanulásra és az alkalmazkodásra. A kutatók különböző módszereket használnak a HMR tanulmányozására, például laboratóriumi kísérleteket, felhasználói értékeléseket és szimulációs modellezést. Mindezek mellett egyre nagyobb hangsúly helyeződik a MI-alapú tanulástámogató rendszerekre is, pl. MI alapú online kurzusok, MI oktatási tananyagok, MI-alapú

teljesítményértékelés és így tovább. Az előadás a HMR fontossága és szerepe mellett az MI-alapú módszertani és technológiai lehetőségeket és megoldásokat mutatja be.

Prof. Dr. Molnár György, Dékán, Óbudai Egyetem Kandó Kálmán Villamosmérnöki Kar. E-mail: <molnar.gyorgy@uni-obuda.hu>.

György Eigner

The Future of Informatics Teaching in the Shadow of Large Language Models – LLMs

The future of informatics teaching is fundamentally changed by AI-driven technologies. As a consequence of today's AI boom more than one framework or system seems to appear on the horizon: assistive systems (co-pilot system types) and AI systems types that are able to generate complete softwares. In the domain of system design the demand according to present trends is shifting from the dominance of development in the direction of integration and control, testing. This requires basic alterations in teaching, too, since engineering-developing reflection will be transferred at least in the field of code expansion to controlling and integrating functions. A further challenge is that the extraordinary source demand of AI-driven systems permeates education, too, a challenge for which as for now we have no solutions. Thus the dominance of the IT infrastructures possessed by great US technological companies will also in the domain of teaching constitute questions we have no simple answers for.

György EIGNER (PhD, Associate Professor, Óbuda University, John von Neumann Faculty of Informatics, Dean; Biomatics and Applied Artificial Intelligence Institute, Head of Institution). E-mail: <eigner.gyorgy@nik.uni-obuda.hu>.

Galambos Péter

A nagy nyelvi modellek után jönnek a nagy motorikus modellek?
Az "Embodied AI" forradalma

Az önreprodukcióra és hasznos dolgok előállítására képes gépek gondolata többek között Neumann-t is foglalkoztatta, sőt eredményei máig meghatározóak. Ma azonban nem az önreprodukción, hanem sokkal inkább az autonóm módon elvégzett hasznos dolgok gépi megvalósításán van a figyelem fókusza. Elég, ha csak a közlekedésre, idősgondozásra, háztartási teendőkre, vagy a gyártásban előforduló feladatokra gondolunk. Ezeken a területeken az elmúlt 10–15 évben lassú, de egyenletes fejlődés figyelhető meg, aminek egyik jellemzője, hogy speciális feladatok mentén, céleszközökben jelennek meg az újdonságok, amelyek nem érnek össze és nem alkotnak egy általános fundamentumot. Ezzel szemben a fizikai közeghez közvetlenül nem kapcsolódó mesterséges intelligencia területek fejlődése felgyorsult és széles körben elérhető, kommersz paradigmaváltó újdonságok jelentek meg 1–2 év leforgása alatt. Ez részben annak köszönhető, hogy a neurális modellek létrehozására fordított szellemi és anyagi erőforrások léptéke ugrásszerűen, nagyságrendekkel megnőtt. Az „alkotó mesterséges intelligencia” (Creative AI) sikerei ma elsősorban a szöveges, vizuális és audio anyagok előállításában érhetők tetten, de mi a helyzet a fizikai világgal? Előadásomban ezt a kérdéskört igyekszem megvilágítani jelenleg folyó kutatások és trendek optikáján keresztül.

Galambos Péter, az Óbudai Egyetem Kutató és Innovációs Központjának főigazgató-helyettese; a Bejczy Antal iRobottechnikai Központ, BARK, igazgatója. E-mail: <mailto:peter.galambos@irob.uni-obuda .hu>.

Nyíri Kristóf

Mesterséges intelligencia, avagy a szellemi elit hanyatlása

Jelen workshop címe, *MI: Robotikától a filozófiáig: A jövő intelligens robotjai – avagy az emberi evolúciós fejlődés mesterséges intelligencia alapokon*, provokatív cím. Filozófiai hátterem nézőpontjából arra érzem magam provokálva, hogy jelezzem: nem a robotok intelligensek, hanem az informatikusok és mérnökök, akik a robotokat építik. Robotok nem lehetnek intelligensek, mivel nincsen tudatuk. Meglehet tökéletesen elvégeznek valamely feladatot, de nincsen tudásuk arról, hogy mit tesznek.

Belevágva előadásom témájába, érvelésemet az önjáró – ha tetszik „intelligens" – autók példájával kezdem. Az ilyen autók *használóira* vonatkozóan van aggodalmam. Ha tényleg hozzászoknak ezekhez a járművekhez, vezetési képességeik szükségképpen sorvadnak – tapasztalatlan autóvezetőkké lesznek. Közelebb a jelenhez, már a GPS használata is csökkenti az autóvezetés tudatosságát. Mondjuk valaki el akar autózni az egyik faluból egy másik, nem túl közeli faluba. Ahelyett, hogy előbb térképre nézne az útvonalat megismerendő, majd intelligensen tájékozódva haladna, folyton a kijelzőt nézi és úgyszólván vakon vezet.

Egy másik, nagyon más – és számomra riasztó – példa a ChatGPT eljövetele. A kilátás az, hogy bárki képes lesz intelligensnek látszó dokumentumok generálására olyan területeken, amelyeket csak kevéssé vagy akár egyáltalán nem ismer. Nem vezet-e ez összeomláshoz a tudományos publikálás, sőt az egész felsőfokú oktatás rendszerében? Hiszen a lektorok vagy egyetemi oktatók számára immár nem elegendő, ha gyanú esetén egyszerűen ránéznek az internetre.

Aligha találnak árulkodó nyomokat. Ám találhatnak nyomokat magában a dokumentumban – téves utalásokat, alaptalan gondolatokat – amennyiben eléggé – széleskörűen és mélyen – műveltek. Ezzel elérkeztem előadásom fő mondanivalójához: új szellemi elitre van szükségünk, olyan elitre, amely kiismeri magát mind Gutenberg, mind Turing világában.

Befejezésképp a tudományos kétnyelvűség dilemmáira szándékozom utalni a nem-angol anyanyelvűek szempontjából, kifejezve reményemet, hogy jelen workshop mintázata éppenséggel e dilemma feloldhatósága felé mutat irányt.

Nyíri Kristóf, MTA r.t., az MTA Kommunikáció- és Médiatudományi Osztályközi Állandó Bizottságának elnöke. E-mail: <nyirik@gmail.com>.

FELTÖLTÖTT ELŐADÁS

Nyíri Kristóf

Mesterséges intelligencia, avagy a szellemi elit hanyatlása

Jelen workshop alcíme – „A jövő intelligens robotjai" – a magam filozófiai háttere felől tekintve: provokatív alcím. Ezt a hátteret életemen át inspiráló hősöm, Ludwig Wittgenstein alakította. Az MI témájával először egy 1985-ben tartott előadásomban foglalkoztam,[1] a *Filozófiai Vizsgálódások* híres sorait idézve: „Dehát egy gép nem tud gondolkodni! – Tapasztalati állítás ez? Nem. Csakis az emberről, és arról, ami hozzá hasonló, mondunk olyat, hogy gondolkodik. … Tekintsd a 'gondolkodni' szót szerszámnak!"[2] Ha a „gondolkodni" szót gépre alkalmazzuk, sugallja Wittgenstein, puszta logikai hibát követünk el. Egy következő előadásban, melyet a *Mesterséges Intelligencia Filozófiai Nézőpontból* című Nemzetközi Nyári Iskolán, Bolzanoban, 1988 júliusában tartottam, ismét visszatértem Wittgenstein-hoz, ezúttal ún. *Kék Könyv*-éből idézve: „a problémát, amely itt fölmerül, a következő kérdéssel lehetne kifejezni: 'Lehetséges-e, hogy egy gép gondolkozzon?' … És a zavar, amelyet ez a kérdés kifejez, igazából nem az, hogy még nem ismerünk olyan gépet, amely alkalmas volna a feladatra. … A zavar inkább abból adódik, hogy a mondat 'A gép gondolkodik (érzékel, kíván)': valahogy értelmetlen."

---

[1] "Wittgensteins Aufhebung der Gestalttheorie", *Akten des Neunten Internationalen Wittgenstein Symposiums*, Wien: Hölder-Pichler-Tempsky, 1985, újranyomtatva *Gefühl und Gefüge* c. kötetemben, Amsterdam: Rodopi, 1986, ld. itt a 202–204. o.-t. Ebben az előadásban hangsúlyosan utaltam Joseph Weizenbaum élesen MI-ellenes könyvére: *Computer Power and Human Reason: From Judgment to Calculation* (1976), valamint a Turing–Wittgenstein kapcsolatra.

[2] Neumer Katalin fordítása.

Előadásom 1989-ben került megjelentetésre, „Wittgenstein és a gépi intelligencia problémája” címmel.[3] Az előadásban mérlegeltem egyfajta lehetséges érvelést, miszerint a számítógépkultúra radikális felemelkedésével, s egyszersmind az életformánkban és nyelvhasználatunkban bekövetkező radikális változással, a gondolkodó gépre történő utalások végső soron mégis értelemmel bírhatnak.[4] Ma teljességgel elvetem ezt az érvelést, az okok az alábbiakban nyilvánvalóvá lesznek. – Bevezetésképp azt állítottam, hogy az „intelligens robotok” kifejezés provokatív. Kifejtve: filozófiai hátterem nézőpontjából arra érzem magam provokálva, hogy jelezzem: nem a robotok intelligensek, hanem az informatikusok és mérnökök, akik a robotokat építik. Robotok nem lehetnek intelligensek, mivel nincsen tudatuk. Meglehet tökéletesen elvégeznek valamely feladatot, de nincsen fogalmuk arról, hogy mit tesznek.

Belevágva előadásom voltaképpeni témájába, érvelésemet az önjáró – ha tetszik „intelligens” – autók példájával kezdem. Az ilyen autók *használóira* vonatkozóan van aggodalmam. Ha tényleg hozzászoknak ezekhez a járművekhez, vezetési képességeik szükségképpen sorvadnak – tapasztalatlan autóvezetőkké lesznek. Közelebb a jelenhez, már a GPS használata is csökkenti az autóvezetés tudatosságát. Mondjuk valaki el akar autózni az egyik faluból egy másik, nem túl közeli faluba. Ahelyett, hogy előbb térképre nézne az útvonalat megismerendő, majd intelligensen tájékozódva haladna, folyton a kijelzőt nézi és úgyszólván vakon vezet.

Egy másik, nagyon más – és számomra riasztó – példa a ChatGPT eljövetele. A kilátás az, hogy bárki képes lesz intelligensnek látszó dokumentumok generálására olyan területeken, amelyeket csak kevéssé vagy akár egyáltalán nem ismer. Nem vezet-e ez összeomláshoz a tudományos publikálás, sőt az egész felsőfokú oktatás rendszerében? Hiszen a lektorok vagy egyetemi oktatók számára immár

---

[3] „Wittgenstein and the Problem of Machine Consciousness”, a *Grazer philosophische Studien*-ben, 33/34. évfolyam.

[4] Ezt az érvelést Otto Neumaier adta elő “A Wittgensteinian View of Artificial Intelligence” c. tanulmányában, megjelent a Rainer Born által szerkesztett *Artificial Intelligence: The Case Against* c. kötetben, London–Sydney: Croom Helm, 1986.

nem elegendő, ha gyanú esetén egyszerűen ránéznek az internetre. Aligha találnak árulkodó nyomokat. Ám találhatnak nyomokat magában a dokumentumban – téves utalásokat, alaptalan gondolatokat – amennyiben eléggé – széleskörűen és mélyen – műveltek. Jelen előadásom absztraktjának megfogalmazása idején, s az azóta eltelt hetek során, a témáról számos összefoglaló esszét publikáltak, hadd említsek csak kettőt: a *Nature* című folyóiratban jelentek meg, az első ez-év szeptember 8-ikán, „Tudományos nyomozók szemfényvesztő ChatGPT-használatra bukkantak közleményekben" címmel, a második október 13-ikán, „A ChatGPT-használat megmutatja, hogy a pályázati rendszer törött" cím alatt. Ezen esszék szerzői nyilvánvalóan kiismerik magukat mind Gutenberg, mind Turing világában: az új szellemi elithez tartoznak. Elithez, mely lázad a ma egyetemi berendezkedése ellen, a publikálj vagy pusztulj el imperatívusza ellen, és persze a pályázati rendszer ellen.[5]

Befejezésképp hadd utaljak jelen workshop egy másik provokatív elemére: a tudományos kétnyelvűség dilemmáira. A kétnyelvűség ma széles körben élesen vitatott kérdéssé vált. Tudósok olyan országokban, ahol az angol idegen nyelv, panaszkodnak, hogy ők ezáltal hátrányos helyzetbe kerülnek a világméretű kutatás és publikálás területén. Másrészt kétségtelen tény, hogy jelenleg az angol a tudomány nemzetközi nyelve. Workshopunk kezdeményezői – közöttük jómagam – azt remélik, hogy időben megosztva először magyar, majd angol nyelvű blokkokat rendezve olyan mintázatot találtak, amely éppenséggel e dilemma feloldhatósága felé mutat irányt.

Nyíri Kristóf, MTA r.t., az MTA Kommunikáció- és Médiatudományi Osztályközi Állandó Bizottságának elnöke. E-mail: <nyirik@gmail.com>.

[5] Ld. ehhez "The Merits of Self-Publication" c. jegyzetemet is.

ENGLISH SECTION

ABSTRACTS

Richard Frohock – Stephen Turner

AI as Symposiast

Since the launch of OpenAI's ChatGPT in December 2022, philosophers have debated whether or not we should view this rapidly evolving technology with fear or excitement. One outcome seems unlikely. Due to the way in which artificial intelligences are trained and operate, the possibility of a future AGI ever being able to synthesize new knowledge beyond fairly narrow limits seems unlikely at best. Barring a revolutionary change in how AI works – one that represents a significant departure from how current cutting-edge LLMs operate – future AIs appear incapable of exceeding their training data only by rearranging what they already "know". To be sure, much conventional philosophical writing does just this: deriving A from B rather than B from A. However, this feature actually makes AI a powerful tool for philosophers, but of a specific kind. The value of AGI within philosophical research comes not from what it might be able to do but from that which it is unable to do. What it is unable to do is provide philosophical truth, but to provide a simulacrum of philosophical truth, not unlike the partners to Socrates in the Platonic Dialogues. By training on not only great but also the more obscure works of philosophy, LLMs can open up philosophers to perspectives that would otherwise have gone overlooked. For this reason, AI can make a wonderful interlocutor for philosophers, opening them up to a broader debate and presenting them with the intellectual challenges necessary to push the discipline's knowledge further. Interactions with AIs like ChatGPT can bring to the forefront

flaws in the established knowledge of the domain, making it easier for the researcher to address these flaws in their own work. But it can also provide arguments that are inadequate in their original context but revelatory if transferred to novel contexts: something the models are not equipped to do, but users can.

Stephen Turner is Distinguished University Professor and Graduate Research Professor at the Department of Philosophy of University of South Florida. He is Director of the Center for Social and Political Thought at the University of South Florida. E-mail: <turner@usf.edu>.

Richard R. Frohock is a graduate student at the University of South Florida. His research is focused on the moral and political implications of expert/ non-expert relationships. He is also interested in the epistemic challenges presented by global problems such as climate change, artificial intelligence, and the COVID-19 pandemic. E-mail: <frohockr@usf.edu>.

András Falus

AI in Medicine; Perspectives and Challenges

The widespread adoption of artificial intelligence (AI), in particular the deep-learning technique, made possible by the unrestricted use of databases, the utilization of big data, significantly increased processing power. With its both challenges and tasks, the medical sector is just beginning to realize an impact of generative AI on at multiple levels. AI supports rapid, accurate laboratory, genetic and image interpretation. Patients get access and follow their own data to enhance health education.

Continuous self revision of AI operation by backward propagation results in increasingly precise performance, accuracy, and speed. The internal layers of ML algorithms' interpretability should be improved for xAI (explained AI) medicine to be successfully implemented.

The robustness, usability, and utility of AI will undoubtedly transform of clinical research and medicine, accompanied presently with hopes and concerns.

Prof. Dr. András Falus, Full Member of the Hungarian Academy of Sciences; Professor Emeritus, Department of Genetics, Cell- and Immunobiology, Semmelweis University, Faculty of Medicine. E-mail: <falus.andras@med.semmelweis-univ.hu>.

Eörs Szathmáry

Robotic Hardware, Robotic Mind: The Major Challenges

We are facing spectacular developments in robotics. Robotic bodies are becoming increasingly flexible and versatile, and AI can contribute a lot to the enhancement of the software governing their actions. It is arguable that no consciousness (at least from a human point of view) can be expected to develop in robots without tight coupling between embodiment and software, and without evolution of some sort linking generations of robots that show hereditary variation. The menace is that the latter can pave the way to selfish robots that may act as competitors to humans, amounting to a new major, unwelcome, evolutionary transition.

Prof. Dr. Eörs Szathmáry, Full Member of the Hungarian Academy of Sciences; Member of the Presidency of the Hungarian Academy of Sciences; Professor of Biology, Institute of Evolution, Centre for Ecological Research; Institute of Biology, Eötvös University; Parmenides Center for the Conceptual Foundations of Science. E-mail: <szathmary.eors@gmail.com>.

Valéria Csépe

Are we Afraid or Anxious about Artificial Intelligence?

The concept and research of artificial intelligence (AI) is not new, so the overwhelming increase in the number of publications, conferences, and enthusiastic or concerned newspaper articles on the issue could be surprising to many. Why now, when generations of natural scientists, mathematicians, and philosophers have been dealing with AI since the 1950s? The answer is seemingly simple; AI has left the research phase, there are hardly any digital technical obstacles to its application, it is attracting general attention, thanks to the use of one decoder type in communication and text creation (ChatGPT). An understandable human response is therefore fear and anxiety, especially since our emotions and thinking have been influenced by the world of AI carriers (e.g. robots) who have been experienced as a threat to human existence since the publication of 20th century sci-fi novels and films.

The presentation presents a few examples of the theoretically endless possibilities of use, selected from two categories. The first includes those whose subject is known, i.e., our fear is based on understood information. The second focuses on those we have little or no information about, so that our scary intuitions have no definable object, i.e., we are anxious. All of this does not only characterize the laymen's behavior, as we researchers are also concerned that AI may surpass the human mind, and with its incredible self-development (deep learning) may even pose a risk. AI, like any other invention, can be a boon or a terror depending on the user's intent. Therefore, one of the urgent tasks of the application of AI is authentic,

scientifically based communication, knowledge dissemination, as well as transparent regulation that also determines safety. It does matter what kind of natural (intellectual and emotional) intelligence applies the AI-based inventions today and tomorrow.

Prof. Dr. Valéria Csépe, Full Member of the Hungarian Academy of Sciences; Member of the Presidency of the Hungarian Academy of Sciences; Chair, Hungarian Accreditation Committee; Hungarian Research Network Brain Imaging Centre, Budapest. E-mail: <csepe.valeria@ttk.hu>.

Gerd Gigerenzer

How to Stay Smart in a Smart World

From self-driving cars and dating apps to ChatGPT, the increasing presence of AI has been widely championed – but there are limitations and risks too. Is more data always a good thing? When do algorithms make better decisions then humans, when not? What is the difference between human and machine intelligence? How to tell realistic expectations about AI from marketing hype and techno-religious faith? Are we sleepwalking into surveillance? To answer these questions is helpful for staying in control in an increasingly automated world. In this talk, I introduce the stable-world principle, the adapt-to-AI principle, and the Russian tank fallacy. We need, now more than ever, to arm ourselves with knowledge about how to make better decisions in a digital age.

Prof. Gerd Gigerenzer is Director of the Harding Center for Risk Literacy at the Max Planck Institute for Human Development Berlin, and since 2020 at the University of Potsdam. For details see: https://www.hardingcenter.de/en/the-harding-center/about. E-mail: <gigerenzer@mpib-berlin.mpg.de>.

James E. Katz

Why Machines Will Rule the World (Mostly)

Will machines rule society? Yes, largely, and perhaps entirely. In fact, they already largely do in some contexts. But first, let's define "rule" since here it could include:

- Making decisions that affect human lives.
- Enforcing laws and regulations.
- Using violence or coercion to achieve its goals.
- Controlling the resources and assets of society.

If "rule" is defined as "making decisions that affect the lives of humans", then it seems we are on the cusp of this situation already. Credit availability, commercial solicitations and offers of college admission are in some cases determined by machines. While humans may set the initial parameters, under machine learning regimes the computers find algorithms that improve on the initial performance parameters in ways that are not understandable to humans. Thus it is the machines that are making decisions affecting the lives of humans. The same applies to enforcing some laws and regulations. Traffic tickets are issued by computer-based systems without human intervention; these can lead to fines and arrests (still yet by humans) if not properly resolved. Roving robots have the ability to disable and kill people; drones have autonomously killed humans who fit an AI-determined profile. There is nothing to prevent these systems from becoming generalized, and thus machine rulers settle questions of life or death.

If “rule” is defined as “controlling the resources and assets of society”, there are philosophical precepts that could lead humans to relinquish decisions about social assets and physical resources to machines. Examples are:

- Commitment to care
- Environmental ethics
- Human flourishing
- Utilitarianism
- Distributive justice
- Rise of transhumanism

After all, even without computers, we as members of society have surrendered many freedoms and much treasure for the collective benefits of government, social welfare, and security. Wouldn’t we continue this trajectory if we were persuaded it was philosophically correct?

James E. Katz, Ph.D., Dr. h.c., is Feld Professor of Emerging Media; Division of Emerging Media Studies, College of Communication, Boston University. E-mail: <katz2020@bu.edu>.

Barry Smith

Why Machines Will Never Rule the World

Is the Singularity nigh? Will our conscious selves one day be embodied in silicon? Will we one day become slaves to computing machines rather than their masters? In our book *Why Machines Will Never Rule the World* (Routledge, 2022), Jobst Landgrebe and I argue that all of these questions must be answered in the negative. AI is and will always remain entirely different from the *general intelligence* (GI) of humans: the one is a logic system, while the other is a complex system of the sort that is ubiquitous in the world of living organisms. It is logically impossible for AI to develop into GI. To assume that it can do so is to fundamentally misunderstand the nature of the human mind.

Barry Smith is Professor in the University at Buffalo. He is the author of more than 700 scientific publications, including 15 authored or edited books, and he is one of the most widely cited living philosophers. E-mail: <ifomis@gmail.com>.

UPLOADED PAPERS

Richard Frohock – Stephen Turner

AI as Symposiast

Since the launch of OpenAI's ChatGPT in December 2022, philosophers have debated whether or not we should view this rapidly evolving technology with fear or excitement. One outcome seems unlikely. Due to the way in which AI models are trained and operate, the possibility of a future AI being able to synthesize new knowledge beyond fairly narrow limits appears more science fiction than genuine possibility. Barring a revolutionary change in how AI works and is trained, future AIs appear incapable of doing much more than rearranging the knowledge of their training data. However, this feature makes AI a powerful tool for philosophers, but of a specific kind. The value of AI within philosophical research comes not from what it might be able to do but from that which it is unable to do. AIs cannot provide philosophical truth, but rather a simulacrum of philosophical truth, not unlike the partners to Socrates in the Platonic Dialogues. By training on not only great but also the more obscure works of philosophy, AI can open up philosophers to perspectives that would otherwise have gone overlooked. For this reason, AI can make a wonderful interlocutor for philosophers, opening them up to a broader debate and presenting them with the intellectual challenges necessary to push the discipline's knowledge further. Interactions with AIs like ChatGPT can bring to the forefront flaws in the established knowledge of the domain, making it easier for researchers to address these flaws in their own work while also providing arguments that are inadequate in their original context but revelatory if transferred to novel contexts: something the models are not equipped to do, but users can.

The history of philosophy is full of invented interlocutors: the Athenian Stranger, Rameau's nephew, largely invented stories about the political life of the Iroquois to name a few. What is striking about these cases is their similarity to the AIs that we think of as potential sources of philosophical knowledge. They are in each case denuded of the features that ordinarily allow us to assess the probity and reliability of a speaker, features of their ongoing conduct that reveal existential contradictions between what they do and say, even when the "doing" in question is a matter of joint attention to a situation or object. These contradictions are a primary but not exclusive source of philosophy. Interlocutors provide an abstracted or depersonalized form of this problem: we expect agreement, but we get questioning or disputation from them, or incomprehension. AI plays this same role, with the same limitations: it abstracts from context or standardizes it, and removes the personal and locally conventional information that we normally use in assessing the statements of others.

The use of invented interlocutors is reflective of the dialectic process, consisting of a movement from thesis and antithesis to synthesis, then to a new anti-thesis to the synthesis, providing a familiar model of philosophical discourse. Indeed, it is common practice amongst philosophers to offer and then refine "straw men" arguments when creating new ideas. A primary reason for this is a failure of self-awareness: we do not know what we are "assuming", in the sense that we do not know whether someone from a different tradition would decline to accept some reasoning or inference we make without an intervening step: an enthymeme, hidden premise, or implicit premise. The interlocutor makes these assumptions more apparent so that they may be addressed by the speaker or the interlocutor herself. The interlocutor acts as a metacognitive device, helping us think about our thinking. But, it has the additional property of being coerced in the sense that to persuade the interlocutor, the speaker must identify and validate the missing or implicit premise. This requires the challenge of the interlocutor. Unchallenged, we are often unaware of the "implicit premises" of our inferences, but we can often invent placeholders that serve the purpose when called upon.

In its current – and likely near future – state AI has the potential to coerce these metacognitive moves better than an imagined interlocutor ever could. Since we are often unaware of our own implicit premises, the interlocutor we imagine is liable to share in at least some of these assumptions. Though she attempts to simulate external positions, she is still a part of our own thought. AI, though bound to the assumptions of its training data, represents a broader range of thinking, thus allowing it to overcome and expose our individual assumptions. When incorporated into our cognitive process, AI can become a part of our thinking that allows us to expand our own metacognitive abilities beyond the limits of an *individual* epistemic standpoint. For future philosophers, it may be that AI training can be extended to include this kind of coerced metacognition, and even to access standardized answers to meta-cognitive questions about implicit questions, allowing it to not only challenge the researcher but also resolve these challenges. Indeed, there may be enough quasi-formal grammatical structure to answer many of these interlocutor questions simply by scanning similar answers that could be found in the training data, exposing the philosophical revelations that were left unnoticed due to the assumptions of the era. If we were to scan in all of the philosophy of ancient Greece with such an AI, we might find that such inventions as the modal realism of David Lewis's possible worlds, but maybe not Schrodinger's Cat. The integration of AI into philosophical research may prevent future revelations from going unnoticed.

This is, of course, speculative. But we can plausibly infer that there is a discrepancy between disembodied AI and embodied thinkers whose thought is facilitated and indeed enabled by habits of mind that they cannot normally articulate, and whose conduct is determined by bodily forces interacting with the mind and one another. The AI model is free of these forces. But this is less a liberation than a limitation. The model may be able to be trained to metacognize and resolve grammatical set-pieces, but it cannot be trained in metacognition over feelings. And from what we know of the human mind, emotion and embodiment is bound up with reasoning and intellectual invention. For this reason, AI does not seem to be a direct threat to

the work of the philosopher. While AI may be able to expand the philosopher's ability to contemplate and invent, it is nonetheless dependent on her for its outputs to hold any philosophical significance. Philosophy is a human activity and therefore must be done within the human context, warts and all, to be meaningful.

Stephen Turner is Distinguished University Professor and Graduate Research Professor at the Department of Philosophy of University of South Florida. He is Director of the Center for Social and Political Thought at the University of South Florida. E-mail: <turner@usf.edu>.

Richard R. Frohock is a graduate student at the University of South Florida. His research is focused on the moral and political implications of expert/non-expert relationships. He is also interested in the epistemic challenges presented by global problems such as climate change, artificial intelligence, and the COVID-19 pandemic. E-mail: <frohockr@usf.edu>.

James E. Katz

Why Machines Will Rule the World (Mostly)

Will machines rule society in the future? In some places they already do. And the breadth of their rule seems destined to increase. Will they come to rule the world? To answer that question, we need first to define the term "rule" in the context of algorithmically enabled computers, sensors, robots and other machines. Here it can be operationally defined as characterized by at least one of the following four activities:

- Making decisions that affect human lives;
- enforcing laws and regulations;
- using violence or coercion to achieve its goals; and
- controlling the resources and assets of society.

Let us take a look at whether one or more of these four conditions currently hold and whether the forces operating within them are likely to expand as a result of the availability of autonomous machine activity.

1. Making Decisions that Affect the Lives of Humans

We are already at this point. For example, credit availability is often determined by machines using complex algorithms that take into account a variety of factors, including employment, credit history, income, and debt levels. Google and Facebook use algorithms to target

ads to users that are likely to be of interest to them; thus, some people may never have an opportunity to learn about certain products.

Beyond intangible consumer decisions, AI is being used in ways that affect humans on a physical level, sometimes with helpful or harmful consequences. Here a prime example is self-driving cars and trucks. In terms of affecting the lives of humans, one cannot formulate a more extreme instance than that of killing a person. And that has already happened in this domain, albeit unintentionally, through the operation of the controlling algorithms. The first recorded example of a self-driving car killing a person occurred in Arizona in 2018. An autonomously equipped Volvo collided with a pedestrian crossing the road. At the time, there was a human in the driver's seat ostensibly acting as a monitor and backup to the AI system. However, she failed to intervene in a timely way, apparently distracted by a gadget. Although not a fatal collision, in October 2023, a self-driving car in San Francisco struck and pinned down a pedestrian. In this instance, the vehicle was a Cruise AV operating entirely autonomously, that is, there was no driver inside the vehicle.

Self-driving cars are likely to become even more widespread and as such our machines will increasingly need to make life-or-death decisions. For example, a self-driving car has to choose between swerving to avoid a pedestrian and colliding with another car, either choice risking a fatality. Even though many possible algorithms could be used to guide these decisions, unexpected situations can arise that require the car to improvise. In such cases, the car would rely on algorithmic interactions to make new judgments about the appropriate course of action. This would make the car not only autonomous but also potentially unpredictable in its behaviour, with direct implications for human lives.

2. Enforcing Laws and Regulations

Cameras operated by AI automatically detect and report traffic violations, such as speeding and red-light running. Traffic tickets are issued by computer-based systems without human intervention; these can lead to fines and arrests (still yet by humans) if not properly resolved. AI systems are used autonomously in the identification process of enforcing environmental regulations. They also make autonomous decisions concerning allowing travelers to enter a country, as may be seen in the automatic gates at international airports. Using facial recognition and documentation, they use algorithms to independently allow a person into a country.

Pressure to use security and regulatory resources efficiently will lead to growing dependence on autonomous policing. As machines become more capable, they will be moving up the ladder of authority and human-handling capabilities. Moreover, we can expect algorithms to continue to make recommendations to judges and probation officers about disposition of criminal cases. As bias gets squeezed out of these systems and they become better at prediction, it may well be that they will be given the responsibility to make the actual decisions about the cases.

3. Using Violence or Coercion to Achieve Its Goals

Roving robots have the ability to patrol areas and engage with any people they encounter. Military robots have the potential to autonomously disable and kill people. Thus far these robots have remained under human control but it’s easy to anticipate their autonomous use during combat. Indeed, in the Ukraine war autonomous drones have apparently engaged and attacked humans. Ukrainian operators have said that their drones can not only independently find, identify and attack Russian military assets, including soldiers, but they have already done so. As drone capabilities improve, there will be increas-

ing reliance on their algorithmic systems to independently make judgments about attacking a human target.

4. Controlling Resources and Assets of the Society

Currently there are autonomous AI systems operating in financial markets. Many major stock markets are run by algorithmic processes the interaction of which is a source of great concern as they may precipitate market crashes. Less dramatically, many firms use AI to manage their supply chains and inventory. With the expectation that these autonomous systems will produce greater benefits for clients and firms, it is to be expected that they would proliferate.

5. Is the Trajectory Destiny?

Humans set the initial parameters for algorithmic machines, but these machines can assign themselves new algorithms under machine learning regimes. This self-improvement process can drive performance beyond initial parameters in ways humans cannot understand. AI-driven machines will solve problems beyond human capacity. In science and health, these machines have discovered unexpected relationships and designed surprising new drugs. Given such success, it is little wonder that decision-makers are now looking for ways to apply AI to societal problems, ranging from education to healthcare and beyond.

In pursuit of a better society, and with a cornucopia of benefits dangled before us, there is seemingly no limit to how much power and control our societies are ultimately willing to cede to AI-driven machines. J. Robert Oppenheimer commented in another context that Edward Teller's proposal to build a hydrogen bomb offered a path that would be too sweet to resist pursuing. The promise of manifold blessings offered by autonomous machines may also prove too sweet to resist.

Parallel to the H-bomb case, there are voices urging us to not pursue the path of greater machine autonomy, that is, "rule". There are profoundly disturbing concerns – both among elites and ordinary people – about a power transfer from the metaphorical hand of humans to those of robot grippers.

Yet these legitimate fears notwithstanding, there is much to be commended about efforts to create a system that can make better decisions than humans themselves. By extending the promise of relieving society's ills, autonomous machines could, step-by-step, forge a utopian society. Or so it might seem at the outset.

Recent history is replete with leaders who initially promised to construct an ideal society but ended as catastrophic failures. Despite their glowing promises, such leaders and their regimes devolved into brutal dictatorships featuring mass killings. Yet despite these lessons, some argue that it is worth trying again to pursue a better world by employing the fantastic power of algorithms, sensors and computers. With machine-learning computers, they say, this time it will be different.

James E. Katz, Ph.D., Dr. h.c., Feld Professor of Emerging Media; Division of Emerging Media Studies, College of Communication, Boston University. E-mail: katz2020@bu.edu.

Barry Smith

Why Machines Will Never Rule the World

In our book *Why Machines Will Never Rule the World*,[1] Jobst Landgrebe and I argue that the efforts of many in the artificial intelligence community to create an artificial *general* intelligence (AGI) are doomed to fail. Here "AGI" is defined as referring to a machine that would exhibit cognitive capacities equivalent to, or even surpassing, those of human beings.

Our argument for the impossibility of such a machine has the following form:

> 1. we analyze the properties of complex systems[2] such as the Earth's weather system or the traffic system of Istanbul,
> 2. we demonstrate that there are severe limits on our ability to predict mathematically the behaviours of systems of this sort,
> 3. we show that these limits then determine also the abilities of computers to make such predictions.

---

[1] Jobst Landgrebe and Barry Smith, *Why Machines Will Never Rule the World:. Artificial Intelligence Without Fear*, Abingdon, UK: Routledge, 2022.

[2] We draw here also on Stefan Thurner, Rudolf Hanel and Peter Klimek, *Introduction to the Theory of Complex Systems*, Oxford University Press, 2018. Note that "complexity" as we use it here is distinct from mere complicatedness. As we shall see, there can be very complicated systems which are yet *simple* – in the book we call them "logic systems" – in the sense that their behavior can be predicted by a machine. That this is so is indeed a trivial matter, since the very act of *providing outputs* by ChatGPT is mathematically a case of *predicting those outputs*.

Our conclusion as to the impossibility of AGI now follows from the fact that all organic systems – including the human neurological system – are complex systems in the sense defined in our book. AGI, however, would require computers which are able to predict with high reliability the behaviours of human beings. This ability is indispensable, for example, if a machine is to participate intelligently in human conversations. And it is the lack of this ability which explains the problems we encounter when dealing over the phone with our bank's computers. For the latter still display a miserable level of performance even after 50 years of attempts by AI engineers to simulate human telephone behaviour. Sophisticated competence in communicating with human beings – for example with controllers of armies, suppliers of munitions, and so forth – would be needed, of course, if machines are to rule over the world.

Our arguments imply that computers will always be restricted to the use of algorithms of the sort that are able to predict only the behaviours of simple systems such as laptops or factory assembly lines. For this reason they will always exhibit what is called "narrow AI", which means that they will always fall short of the *general* intelligence of human beings.

Enter ChatGPT

Or so, when our book was published in September 2022, we had claimed.

But then, in November of 2022, a new and revolutionary kind of AI was unleashed onto the world in the form of "ChatGPT". ChatGPT was different. It represented the first example of an AI system that could be accessed easily by normal human beings and provide hours of annoyance-free stimulation. In addition, it was able to provide a range of different kinds of services that can be of value, for example, to commercial organizations. ChatGPT was also, it seemed,

an example of *general* AI, since it could respond, in its way, to prompts on every subject under the sun.[3]

Very soon, however, the growing community of users of ChatGPT became aware of certain unanticipated behaviours, called "hallucinations", on the part of its algorithm. For the latter is apparently statistically primed to avoid phrases like "I don't know" in its outputs. Instead it will oftentimes invent its own answers to questions where it did not know the answer. In short, it would exhibit behaviour which, if committed by a human, would be referred to as "telling lies".[4]

How ChatGPT Works

To understand what is going on here, we must recognize, first, that ChatGPT does not in fact *know* anything. When it gives a response R to a prompt P this is not because it knows that some proposition R is true.

ChatGPT is a set of algorithms in software form, built on the basis of the GPT probabilistic model of the order of the symbol sequences produced, for example, in conversations or in question-and-answer sessions.

When we enter some prompt P, therefore, then what happens is that the ChatGPT algorithm predicts that, given the data that was used as its training set, a given sequence of tokens (roughly: syl-

---

[3] Blaise Agüera y Arcas and Peter Norvig, "Artificial General Intelligence Is Already Here", *Noema Magazine*, October 10, 2023, https://www.noemamag.com/artificial-general-intelligence-is-already-here/.

[4] The conversation displayed here: https://buffalo.box.com/v/ChatGPT-getting-worse provides an example of this phenomenon, illustrating how ChatGPT is of the opinion that there are three philosophers by the name of J. K. (or J. C.) Nyíri. One, a Jozef (dead since 2018), being the father to the next, a Janos, still alive and "known for his work in logic, philosophy of science, and philosophy of language"; the third, a certain "J. Christoph Nyíri", is said to be an emeritus professor of philosophy in Paris and an expert on the topic of heritagization.

lables) R is (roughly) the *next most likely sequence of tokens* given P as starting point.[5]

We must recognize, second, that ChatGPT functions always only by drawing on the specific (and admittedly very large) body of data upon which it was trained. This implies that ChatGPT, too, is an example of narrow AI. This is because its algorithm does not relate to the real world in which we live and to the many, many, overlapping and ever evolving complex systems out of which it is composed. Rather, it relates to a certain abstract simulacrum of a world, a simulacrum that is exactly specified by the large but finite set of data upon which the algorithm was trained, in something like the way in which the world of a video game is specified by the game's software. For ChatGPT this training data was defined (roughly) by the contents of the internet on some given day in the past.[6]

The algorithm, which was defined through a process of training on the basis of data available in 2021, is a mathematical function which takes as inputs binary vectors encoding prompts P, and outputs binary vectors encoding responses R. We can think of this function as a very, very long polynomial equation (with some 1.5 billion parameters). Hence the broad reach of topics upon which it can provide responses to prompts. At the same time, however, the mathematical abilities encapsulated in this equation are still very simple – since the equation, like all computable algorithms, needs to be capable of being solved by using only the very simple mathematics of a Turing machine.[7]

---

[5] Note that "predict" here means that the algorithm issues R as output given P as input.

[6] Hence its use of the modifier "As of my last knowledge update in September 2021" to justify its not providing answers to questions pertaining to more recent events.

[7] ChatGPT is in this respect comparable to Google Translate. The latter can be applied to many languages used to cover many, many topics. Yet Google Translate is nonetheless an example of narrow AI. Like ChatGPT it works only for very simple and unchanging worlds, which are defined by the data upon which any given version of the software was trained.

Conclusion

We can now understand why, as the material referenced in footnote 4 demonstrates, ChatGPT yields such peculiar results when questions are raised about persons with names like “J. K. Nyíri”, even as it does so well when questions are raised about, say, J. K. Rowling. This is because the internet is rich with data about all things Harry Potter, and the algorithm performs very well when predicting the (next most likely) answers to questions about topics about which it has large amounts of data. It makes such a mess of itself, in contrast, when addressing the topic of a “J. K. Nyíri”, because the world of the internet is to such an overwhelming degree an Anglosaxophone, rather than a Magyarophone, world.

We note, finally, that things will not get better, from the point of view of AGI, even if such “hallucination” problems can be solved. If the GPTs of the future can reduce the degree to which they generate hallucinations, it will still remain the case that the world in relation to which each successive release of GPT provides its responses is an unchanging logic-system-defined world. Thus it is not anything like the world in which, over millions of years, human intelligence has evolved.

Barry Smith is Professor in the University at Buffalo. He is the author of more than 700 scientific publications, including 15 authored or edited books, and he is one of the most widely cited living philosophers. E-mail: ifomis@gmail.com.

ADDENDA / KIEGÉSZÍTÉSEK

Csépe Valéria

Félünk vagy szorongunk a Mesterséges Intelligenciától?

A mesterséges intelligencia (MI) koncepciója és kutatása nem új, ezért sokak számára elvileg meglepő lehetne a kérdésről megjelenő publikációk, konferenciák, a lelkes vagy aggódó újságcikkek számának mindent elsöprő növekedése. Miért most, ha már az 1950-es évektől természettudósok, matematikusok, filozófusok generációi foglalkoztak az MI-vel? A válasz látszólag egyszerű; az MI kilépett a kutatási fázisból, alkalmazásának alig vannak digitális technikai akadályai, általános figyelem övezi, köszönhetően az egyik dekóder típus alkalmazásának a kommunikációban, szövegalkotásban (ChatGPT). Érthető emberi válasz ezért a félelem és szorongás, különösen, hogy érzelmeinkre, gondolkodásunkra a 20. századi sci-fi (tudományos-fantasztikus) regények és filmek megjelenése óta hat az emberi lét fenyegetéseként megélt MI hordozók világa (pl. robotok).

Az előadás néhány példát mutat be az elméletileg végtelen felhasználási lehetőségek közül, mégpedig két szempont alapján válogatva. Az első kategóriába azok tartoznak, amelyeknek tárgya ismert, azaz félelmünk megértett információra épül. A második körben azokról a kihívásokról lesz szó, amelyekre vonatkozóan információnk kevés vagy nincs, ezért rossz megérzéseinknek nincs meghatározható tárgya, azaz szorongunk. Mindez nem csupán az átlagembert jellemzi, hiszen bennünket kutatókat is aggaszt, hogy az MI felülmúlhatja az emberi elmét, s hihetetlen önfejlődésével (mély tanulás) akár kockázatot is jelenthet. Az MI a találmányok sorában ugyanolyan, mint bármely korábbi, a felhasználó szándéka szerint lehet áldásos vagy rémisztő. Ezért az MI alkalmazásának egyik

sürgető feladata a hiteles, tudományosan megalapozott kommunikáció és ismeretterjesztés, valamint a biztonságot is meghatározó átlátható szabályozás. Nem mindegy, milyen természetes (értelmi és érzelmi) intelligencia határozza meg az MI-alapú alkalmazásokat ma és holnap.

Prof. Dr. Csépe Valéria, az MTA rendes tagja; az MTA Elnökségének tagja; a Magyar Akkreditációs Bizottság elnöke; Nemzeti Kutatási Hálózat Agyi Képalkotó Központ, Budapest. E-mail: <csepe.valeria@ttk.hu>.

Gerd Gigerenzer

Hogyan maradjunk okosak egy okos világban

Az előadás a szerző 2022-es sikerkönyvének – *How to Stay Smart in a Smart World: Why Human Intelligence Still Beats Algorithms* – rövid összefoglalása.

Önjáró autóktól és társkereső applikációktól a ChatGPT-ig, az MI erősödő jelenlétét széles körben ünneplik – de látnunk kell a határokat és rizikókat is. Vajon az egyre több adat mindig hasznos? Mikor hoznak algoritmusok jobb döntéseket, mint az emberek, és mikor nem? Mi a különbség az emberi és a gépi intelligencia között? Hogyan válasszuk el az MI-t illető realisztikus várakzásokat az üzleti felhajtástól és a technovallásos hittől? Vajon alvajáróként haladunk az általános megfigyelés világába? Az ezen kérdésekre adott válaszok segítségünkre lesznek abban, hogy továbbra is ellenőrzésünk alatt tartsuk fokozódóan automatizált világunkat. Az előadás bevezeti a stabilvilág ("stable-world") elvet, az alkalmazkodjunk-az-MI-hez elvet, és elemzi az orosz tank tévedést ("Russian tank fallacy" – a kifejezés egy USA MI projektre vonatkozik, amely az orosz és amerikai tankokat volt hivatva egymástól megkülönböztetni, de a napos időben / felhős időben eltérő mintavétel félrevezető volta miatt a való világban nem működött). Ma inkább, mint bármikor, speciális tudásra van szükségünk ahhoz, hogy a digitális világban jobb döntéseket hozzunk.

Az előadás angolul került megtartásra.

Prof. Gerd Gigerenzer a Harding-Zentrum für Risikokompetenz (Max-Planck-Institut für Bildungsforschung, Berlin) igazgatója, és

2020 óta a Potsdami Egyetemen tanít. A részleteket ld.: https://www.hardingcenter.de/en/the-harding-center/about. E-mail: <gigerenzer@mpib-berlin.mpg.de>.

György Molnár

The Development of Human–Machine Interactions – AI-based Innovative Supportive Learning Systems and Learning Environments in Education

During the past few decades in the field of artificial intelligence there occurred significant steps toward developing devices and robots that are able to interact with humans in various contexts. To such robots belong the humanoid or social ones, which are used in many environments, from education through medicine to entertainment and research. The effectivity of these intelligent devices however greatly depends on their ability to connect in a smart and attractive way to people. Human–machine interaction (HMR) is an interdisciplinary research domain, the aim of which is to achieve a deeper understanding of the communication, cooperation and connections between people and machines. In recent years the results of HMR research have been greatly enriched in dimensions like perception, cognition, navigation, and communication technologies, enhancing human–machine collaboration. One of the main tasks of human–machine interaction is to ensure a smooth cooperation between people and machines in order to achieve various tasks. HMR research focuses on various problems, for example on social interactions, collaboration, learning, and adaption. Researchers apply various methods in studying HMR, such as laboratory experiments, user evaluations and simulation modelling. Besides these there is a growing emphasis on AI-based Supportive Learning Systems, e.g. AI-based online courses, AI educational materials, AI-based performance evaluation, etc. This talk will point on

the one hand to the importance and role of HMR, and on the other to AI-based methodological and technological possibilities and solutions.

Prof. Dr. György Molnár, Dean, Kandó Kálmán Faculty of Electrical Engineering, Óbuda University. E-mail: <molnar.gyorgy@uni-obuda.hu>.

György Eigner

The Future of Informatics Teaching in the Shadow of Large Language Models – LLMs

The future of informatics teaching is fundamentally changed by AI-driven technologies. As a consequence of today's AI boom more than one framework or system seems to appear on the horizon: assistive systems (co-pilot system types) and AI systems types that are able to generate complete softwares. In the domain of system design the demand according to present trends is shifting from the dominance of development in the direction of integration and control, testing. This requires basic alterations in teaching, too, since engineering-developing reflection will be transferred at least in the field of code expansion to controlling and integrating functions. A further challenge is that the extraordinary source demand of AI-driven systems permeates education, too, a challenge for which as for now we have no solutions. Thus the dominance of the IT infrastructures possessed by great US technological companies will also in the domain of teaching constitute questions we have no simple answers for.

György EIGNER (PhD, Associate Professor, Óbuda University, John von Neumann Faculty of Informatics, Dean; Biomatics and Applied Artificial Intelligence Institute, Head of Institution). E-mail: <eigner.gyorgy@nik.uni-obuda.hu>.

Kristóf Nyíri

Artificial Intelligence and the Demise of the Intellectual Elite

The subtitle of this conference – "The Intelligent Robots of the Future" – is, seen from the point of view of my own philosophical position, provocative. This position is inspired by my life-long hero Ludwig Wittgenstein. I have first dealt with the topic of AI in a talk given in 1985,[1] quoting the famous *Philosophical Investigations* lines: "a machine surely cannot think! – Is that an empirical statement? No. We only say of a human being and what is like one that it thinks. … Look at the word 'to think' as a tool." Applying the word "to think" to a machine, Wittgenstein implies, is simply to commit a grammatical mistake. In a talk I gave at the International Summer School *Philosophical Perspectives on Artificial Intelligence* at Bolzano, in July 1988, I again came back to Wittgenstein, this time quoting from his so-called *Blue Book*: "the problem here arises which could be expressed by the question: 'Is it possible for a machine to think?' ... And the trouble which is expressed in this question is not really that we don't yet know a machine which could do the job. ... The trouble is rather that the sentence, 'A machine thinks (perceives, wishes)': seems somehow nonsensical." This talk was published in 1989 under

[1] "Wittgensteins Aufhebung der Gestalttheorie", in *Akten des Neunten Internationalen Wittgenstein Symposiums*, Wien: Hölder-Pichler-Tempsky, 1985, reprinted in my volume *Gefühl und Gefüge*, Amsterdam: Rodopi, 1986, see here pp. 202–204. In this talk I have referred with emphasis to Joseph Weizenbaum's markedly anti-AI book *Computer Power and Human Reason: From Judgment to Calculation* (1976), as well as to the Turing–Wittgenstein connection.

the title "Wittgenstein and the Problem of Machine Consciousness".[2] In the talk I contemplated a possible argument to the effect that with the radical rise of computer culture, and thus with a radical change of our form of life and our use of language, references to a thinking machine might make a sense after all.[3] Today I absolutely reject this argument, for reasons that should become clear below. By way of introduction I suggested that the phrase "intelligent robots" is provocative. To explicate: I am provoked to say that intelligent are the people who build and develop robots, but robots themselves cannot be intelligent, since they lack the capacity for thinking. They might perform tasks perfectly, but they have no knowledge of what they do.

Coming now to the actual topic of this paper, I will begin my argument with the example of self-driving cars – if you like "intelligent" cars. My problem is with the *users* of those cars. If they get really accustomed to such vehicles, their driving skills will necessarily decrease – they will become inexperienced drivers. Nearer to the present, already the use of GPS lessens drivers' awareness of what they are doing. Say you want to drive from a village to a not-too-nearby other village. Instead of familiarizing yourself with the route by first looking at a map, and then navigate intelligently, you keep looking at the display and drive so to speak blindly.

Another, very different – and for me frightening – example is the arrival of ChatGPT. The prospect is that one can generate intelligent-looking documents in fields one has little or no knowledge of. Will this not bring about the breakdown of the system of scientific publication, and indeed of the higher education system? Since for reviewers, or for university teachers, receiving papers from their peers or students, this will not be the case of simply checking the internet if suspicions arise. They might well find no clues at all. But they might find clues in the document itself – incorrect references, baseless ideas – if they are sufficiently and indeed widely and deeply educated. At

---

[2] In the journal *Grazer philosophische Studien*, vol. 33/34.

[3] This argument was presented by Otto Neumaier in his paper "A Wittgensteinian View of Artificial Intelligence", in Rainer Born (ed.), *Artificial Intelligence: The Case Against*, London–Sydney: Croom Helm, 1986.

the time I was formulating the abstract for the present paper, and during the weeks that have since passed, a number of summary essays appeared on the topic, let me mention just two, both published in the journal *Nature*, the first one on September 8 this year, under the title "Scientific Sleuths Spot Dishonest ChatGPT Use in Papers", the second on October 13, entitled "ChatGPT Use Shows that the Grant-Application System Is Broken". The authors of these essays are, obviously, well-trained both in the worlds of Gutenberg and Turing, they belong to a new intellectual elite. This is an elite that revolts against today's university establishment, against the publish or perish imperative, and of course the grant-application system.[4]

By way of conclusion let me point to another challenge the present conference poses: that of bilingualism. Bilingualism is becoming in many places a hotly discussed issue. Scholars in countries where English is a foreign language complain that they are thereby disadvantaged in world-wide research and publication. On the other hand it is a plain fact that today English is the international language of science. The initiators of this conference – amongst them myself – hope that by organizing sessions first in Hungarian and then in English they have found a pattern which might bring us closer to resolving the dilemma.

Kristóf Nyíri, PhD, Dr. h.c., is a retired philosophy professor. He is Full Member of the Hungarian Academy of Sciences, and Chair of the Committee for Communication and Media Theory, Hungarian Academy of Sciences. E-mail: nyirik@gmail.com.

[4] See on this also my note "The Merits of Self-Publication".

MI: Robotikától a filozófiáig. A jövő intelligens robotjai – avagy az emberi evolúciós fejlődés mesterséges intelligencia alapokon.

AI: From Robotics to Philosophy. The Intelligent Robots of the Future – Or Human Evolutionary Development Based on AI Foundations.

A Magyar Tudományos Akadémia és az Óbudai Egyetem által rendezett workshop anyagai, szerk. Molnár György és Nyíri Kristóf.

Proceedings of a Hungarian Academy of Sciences / Óbuda University workshop, edited by György Molnár and Kristóf Nyíri.

Prof. Dr. Molnár György, Dékán, Óbudai Egyetem Kandó Kálmán Villamosmérnöki Kar. — Prof. Dr. György MOLNÁR, Dean, Kandó Kálmán Faculty of Electrical Engineering, Óbuda University.

Nyíri Kristóf, MTA r.t., az MTA Kommunikáció- és Médiatudományi Osztályközi Állandó Bizottságának elnöke. — Kristóf NYÍRI, PhD, Dr. h.c., Member of the Hungarian Academy of Sciences; Chair, Committee for Communication and Media Theory, Hungarian Academy of Sciences; Research Professor at Óbuda University.