

Responsible AI

29th SEP Conference 2023

PRAGUE
6TH OF NOVEMBER 2023



UNIVERSITY OF
ECONOMICS AND
BUSINESS PRAGUE



CZECH TECHNICAL
UNIVERSITY IN
PRAGUE



SILESIA
UNIVERSITY
IN OPAVA

Venue: Prague University of Economics and Business,
W. Churchill square 1938/4, Prague 3

Conference Date: November 6th 2023

EMAIL: DEL@VSE.CZ
WEBPAGE: DEL.VSE.CZ



SYSTEM AND ETHICAL APPROACHES 2023 PROCEEDINGS

29th conference

Organizers:

Department of System Analysis
Prague University of Economics and Business

Institute of Computer Science, Faculty of Philosophy and
Science in Opava, Silesian University

Editors:

Richard Antonín Novák
Tomáš Sigmund
Anna Novotná

Abstracts proceedings

Publisher:

Oeconomica Publishing House

systemsapproaches.vse.cz

2023

Programme Committee:
Ing. Richard A. Novák, Ph.D.
Mgr. Ing. Tomáš Sigmund, Ph.D.
PhDr. Ing. Antonín Pavlíček, Ph.D.
Ing. Jiří Korčák
Mgr. Anna Novotná, Ph.D.
Mgr. Kamil Matula, Ph.D.

Design and technical editor: Anna Novotná

**Proceedings of the Conference System and Ethical
Approaches 2023, Responsible AI, technology for good!,
6th of November 2023**

T A
C R

This project is co-financed from the state budget by the Technology agency of the Czech Republic under the TAČR Doprava 2020+ Programme.

1st Edition

2023

ISBN 978-80-245-2508-2

© Oeconomica Publishing House

Systems and Ethical Approaches is licensed under Creative Commons Attribution-ShareAlike 4.0 International License.



Content

| | |
|--|----|
| Richard Antonín Novák | 1 |
| Responsible AI: Technology insight! | |
| Tomáš Sigmund | 3 |
| Automation and Autonomy | |
| Tomáš Tichý, J. Brož, P. Máša, R. A. Novák, J. Štefan and T. Šmerda | 5 |
| Predictive Maintenance in Telematics | |
| Anna Novotná and K. Matula | 9 |
| How AI affects information society | |
| Hynek Trojánek | 13 |
| The future is analogue? | |
| Jiří Korčák | 15 |
| The Power of Human Communities vs Facebook Algorithms | |
| Closing word of Tomáš Sigmund | 17 |

Preface

The current era is riding the hype wave associated with artificial intelligence (AI). In my view, the life of a scientist should not be about hype, but about honesty and also about the fact that every craft, like AI, can be done well or badly. Because craft is something you have to do repeatedly and in depth to be really good at it, and superficiality or shortcuts are not conducive to it.

The field of AI, which many consider just a craft, has currently become not only a trend that is driving society, but also an intersection of multiple disciplines that we are trying to address in a balanced way in our conference. In fact, we do not consider AI to be the domain of computer science and technically educated people only, but we see the importance of the humanities, which should be more involved, especially in the design phase, in defining the goals of AI projects and evaluating their impact.

Responsible AI (rAI), is by many definitions, the practice of designing, developing, and deploying AI with good intentions for both the data scientists who design AI, the users of AI applications, and the broader society that is often subliminally affected by AI. Good practice in rAI then involves the interplay of multiple disciplines such as computer and mathematical sciences, sociology, psychology, philosophy, law and many others.

If, like us, you are interested in the balanced perspectives of data scientists, philosophers, sociologists, lawyers, statisticians, and economists, check out the proceedings of the 29th edition of our traditionally formatted Systems and Ethical Approaches (SEP) 2023 conference. We have attempted here to summarize the conference, i.e., how different professional groups view the issue of Responsible AI.

On behalf of the conference preparatory team.

Ing. Richard A. Novák, Ph.D.

Responsible AI: Technology insight!

Richard Antonín Novák

DEPARTMENT OF SYSTEM ANALYSIS, FACULTY OF INFORMATICS AND STATISTICS, PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

richard.novak@vse.cz

KEYWORDS:

AI – Explainable AI – LIME – Responsible AI – Supervised learning – Reinforcement learning – Unsupervised Learning

ABSTRACT:

The paper presents a comprehensive insight into Responsible AI with a focus on the technological perspective set in a broader context. Within the article, the author applies a multi-disciplinary approach involving the overlap of computing science into related humanities disciplines such as philosophy and sociology in particular.

In the first part, the article describes the roles and motivations of actors in individual data science projects, and emphasizes that rather than talking about ethical AI, we can talk about the ethics of individual development teams working on AI projects. In the introductory section, the paper addresses a number of external factors of data science projects such as the different regulation of AI in the US, EU and China. This, according to the author, besides the specific cultural context, also has a practical impact on the different speed, safety and excellence associated with AI, where Europe currently lags significantly behind its competitors.

In the second part, the paper provides an overview and categorization of the most common data structures such as numeric, categorical, ordinal, unstructured media data among others. Subsequently describes beyond the CRISP approach what impact and limitations the described structures have on their possible further processing. In depth, the author focuses on methods from the area of supervised, unsupervised, reinforcement learning. It then presents typical cases of deployment on specific commercial tasks such as risk, fraud management, recommendation engines, image recognition and voice generation among others.

The third part is devoted to Large Language Models and a special category of foundation models. The author directly follows up the principal introduction of different types of neural networks and deep learning with a discussion of Explainable AI (xAI). This chapter of the article is supplemented with concrete examples of application of LIME method to more complex AI projects, which often represent black boxes for end users.

In the final section, the article summarizes the positives and negatives that AI and its applications bring to society. With the positives most often touching on increased revenue, reduced costs and sources of innovation. The negative aspects of AI then include, for example, enabling the division of society, manipulation, fake news, and uncontrolled influence on the labor market. The concluding paragraph names in a structured way the increasing complexity of AI in particular, which requires the establishment of new approaches to the responsible development of AI beyond the previously known concepts from various commercial companies or the Open Data Institute.

ACKNOWLEDGEMENTS

The presentation was prepared with the support of TA CR under the numbers CK04000082, CK04000109 and FW03010458.

Automation and Autonomy

Tomáš Sigmund

DEPARTMENT OF SYSTEM ANALYSIS, FACULTY OF INFORMATICS AND STATISTICS, PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

sigmund@vse.cz

KEYWORDS:

Artificial intelligence – Ethics – Autonomy – Automation

ABSTRACT:

Currently, artificial intelligence (AI) is becoming a key element shaping our work and personal lives. One of the most significant aspects of AI is its impact on autonomy, both in a professional and personal context. It is important to recognize that increasing dependence on AI can have negative impacts on human abilities and skills.

The automation of routine and repetitive tasks allows people to focus on more complex and creative work. This results in increased productivity and efficiency, but also changes the nature of the work people perform. In manufacturing and logistics, robots are taking over many tasks previously performed by humans, leading to questions about the future role of human workers. This dependence on AI can lead to the degeneration of certain manual and cognitive skills, as people are no longer required to use these skills regularly.

Education is another area where AI is changing the rules of the game. Personalized AI-driven educational platforms can tailor curricula to individual students, which may affect the role of

teachers and students' autonomy in the educational process. Here, there is a risk that excessive reliance on technology can weaken critical thinking and analytical abilities of students.

In transportation, autonomous vehicles promise a revolution in how we move. While this can lead to safer and more efficient transportation, it also raises questions about job loss for drivers and the autonomy of individuals in choosing transportation methods. Additionally, with the increasing use of autonomous vehicles, there could be a decline in driving skills among the wider population.

AI is also changing how we interact with technology on a personal level. With AI's growing ability to adapt to our preferences and behavior, comfort increases, but questions also arise about how much we control technology that is becoming increasingly autonomous. This dependence on AI can lead to a weakening of social skills and the ability to make independent decisions.

In the context of work and autonomy, it is important that technology is integrated properly to respect human capabilities, needs, and values. It is essential to find a balance between using AI to increase efficiency and maintaining human control and decision-making. For example, in medicine, AI can assist with diagnostics, but human intervention is necessary for communication with patients and their families.

In conclusion, AI has the potential to significantly improve our productivity and efficiency, but it's important to realize that this progress comes with challenges related to autonomy and control. As AI becomes increasingly sophisticated, it is crucial that its integration into our lives is conducted with consideration for ethical and social impacts, and ensures that it does not lead to the degeneration of human abilities and skills.

ACKNOWLEDGEMENTS

The presentation was prepared with the support of TA CR under the numbers CK0400082, CK04000109 and FW03010458.

Predictive Maintenance in Telematics

Tomáš Tichý, Jiří Brož, Petr Máša,
Richard Antonín Novák, Jiří Štefan and
Tomáš Šmerda

DEPARTMENT OF TRANSPORT TELEMATICS, FACULTY OF TRANSPORTATION SCIENCES, CZECH TECHNICAL UNIVERSITY IN PRAGUE

tomas.tichy@cvut.cz

jiri.broz@cvut.cz

tomas.smerda@cvut.cz

DEPARTMENT OF INFORMATION AND KNOWLEDGE ENGINEERING, FACULTY OF INFORMATICS AND STATISTICS, PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

petr.masa@vse.cz

DEPARTMENT OF SYSTEM ANALYSIS, FACULTY OF INFORMATICS AND STATISTICS, PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

richard.novak@vse.cz

KEYWORDS:

Telematic systems – Maintenance – Prediction – Artificial intelligence – Diagnostics

ABSTRACT:

This presentation focuses on the application of predictive diagnostics and maintenance of technological equipment and systems that are applied in traffic telematics systems. Examples of such systems include light-controlled intersections, tu-

nnel systems, camera systems and other systems that can also be called intelligent transport systems (ITS). The requirement of administrators of transport and technological infrastructure is not only to ensure the correct functioning and operation of the equipment in the form of maintenance, but also its gradual renewal. Furthermore, last but not least, optimization of operating costs and service costs of ITS equipment.

Intelligent Transport Systems are often distributed over larger areas, and therefore the cost of servicing and renewing equipment is a significant part of the financial cost of deploying ITS systems. These conditions just place great demands on obtaining the necessary data and balancing also other parameters such as reliability, lifetime, failure intensity, life cycle, etc. To determine such parameters, diagnostic data from the ITS equipment is required and predictive diagnostics is needed to ensure the reliability of data.

This is even more important for the case of IT systems implemented in transport that are considered as part of critical infrastructure of the state. The state aim is to operate efficient and reliable maintenance with targeted service interventions. Maintenance requires well-defined documentation, which is prepared and applied in a BIM approach and total digitalization, which greatly assists the orientation of ITS and its technological systems. The goal is to analyze the data obtained, and to combine many parameters to get deeper insights. Furthermore, a methodical approach applied to use of predictive maintenance of ITS technologies is currently often requested to improve the life cycle of the equipment, including the prediction of future states of the equipment.

The presentation contains examples of potential failures that are addressed in operation with an impact on the entire predictive maintenance system from a specific telematics system, which is the tunnel system. Data transfer, redistribution including information from systems and provision of information on equipment status and subsequent life cycle calculation for selected tunnel system equipment is shown here. We believe it

will form the basis for new approaches to ensure secure data transfer, i.e. cyber security and other new approaches where continuity and application of new trends in predictive diagnostics is indicated.

ACKNOWLEDGEMENTS

The presentation was prepared with the support of TA CR under the numbers CK04000082, CK04000109 and FW03010458.

How AI affects information society

Anna Novotná and K. Matula

INSTITUTE OF COMPUTER SCIENCE, FACULTY OF PHILOSOPHY AND SCIENCE IN
OPAVA, SILESIAN UNIVERSITY

anna.novotna@fpf.slu.cz

kamil.matula@fpf.slu.cz

KEYWORDS:

Artificial intelligence – Information behavior – Information overload – Information retrieval – Social media

ABSTRACT:

The topic of artificial intelligence is very popular today. It is expected to take over much of the control of common manufacturing processes, code design, teaching parts, and most mechanically performed operations in the near future. At the same time, it is little known that it already controls our information flows and much of the information curation. AI-powered search engines and recommendation systems enhance the accessibility of information, environment. As a society, we have long used it to sort information, whether in spam filters or in search engine and social networking algorithms. We know that while we need its functionality, it can also have a detrimental effect on the individual and their information behavior, which manifests itself not only physically but also psychologically. In this paper, we reflect on how it actually helps us and how much it potentially harms us. In our paper, we discuss topics that develop the work with information using artificial intelligence.

Firstly, AI is used to access information, allowing users to quickly find it based on their preferences, improving the user experience. A good example is BING from Microsoft or BARD from Google. Filtering and aggregating information using AI helps users avoid information overload. An interesting example from the scientific community is also SciSpace. The negative effect of this form of AI use is disconnection from original sources and problems with engagement with our own information literacy. AI also curates our sources. The problem is that we do not know what the information space looks like. We do not know how precise the answer is or if it does not hallucinate. We only know what AI tells us.

AI also does not apply ethical rules, it does not know what is potentially harmful. AI is taught on a certain data set, and it can become statistically distorted. AI algorithms can mirror unevenly distributed data, which can lead to discrimination against certain groups in decision-making processes, promoting social inequalities, negative behavior, etc.

Other AI tools are used for personalization, for example in social media. It makes the environment more relevant and encouraging to engage. Such personalisation can support users' existing beliefs or expose them to different perspectives. Way of information is organized differs because AI looks not only at the context of the query but also at user preferences, interactions, previous searches, etc. So AI prepares a „tailored“ result for the user.

The problem with personalisation is that it can easily fuel radicalisation. Not only does AI require collecting and using a lot of personal data about users, it also tracks their activity and suggests more and more engaging and radical content. It can draw users into information bubbles and echo chambers, and it prefers advertising profit over balanced information income. This is called attention economy, more radical content will keep users in front of their screens longer.

AI is also used for information verification, for example, by comparing the quality and authenticity of textual, visual, or

audiovisual information. A check can be performed to see if the text is written by AI (for example Grammarly), or generated (for example by Midjourney), or if the information is factual.

The problem is, in this case, accuracy. Even with the best and paid AI tools, the accuracy of 100% cannot be reached. First of all, if you put textual information through more than one AI tool, the accuracy will decrease to a minimum. If the language of the output is not English, it is nearly impossible to reveal plagiarism. The accuracy of the AI visual verification tool also depends greatly on the data used for this task. Accuracy varies with representation of race and gender (etc.) in a data set.

AI is often used for efficient decision making. People have difficulty making decisions in situations that are complex, novel, or in which they have little information. They are prone to mistakes, being swayed by external influences, and they are not able to study „all the data“. Thus, consideration is given to the use of AI in cases where the decision should be impartial and based on data or evidence.

AI can work with large amounts of data and is most applicable when the outcome is clearly defined and measurable. For example it enables businesses to make instant decisions based on quick analysis. Today some applications using natural language processing perform social computing and opinion mining on data from social media. They allow businesses to understand how customers interact with their brand, what words they use, and what things are most appealing to them. This improves the effectiveness of the marketing strategy. At the same time, a better understanding of customer needs and satisfaction is gained.

On the other hand, the error rate and fairness of training data is directly reflected in the quality of AI decisions and the quality of recommendations. People often do not reflect this, and it can lead to erroneous decisions. For example, AI tools used in HR processes can underscore applications not using specific words expected by HR managers. AI in HR tools are not designed to adapt accordingly, they are designed to score accura-

tely. So in case an applicant did not get enough experience in the last few years (even though he were, for example, ill or on parental leave), he is excluded from the „yes“ pile.

Today, applications built using natural language processing over large datasets are widely used. They help with writing grant applications, scientific articles, and term papers, essays, etc. Those who have access to AI technologies and can use AI get better results because the training data are collected from the best. This is why scientists are starting to talk about the digital divide, especially in the education environment. Because basic education is aimed at the largest group of the population. What happens if the educated do not have access to AI technology and do not get the opportunity to have better outcomes? What will become of their life chances?

These are the questions that interest us as individuals participating in the infosphere. What kind of element will AI bring into these areas? To which extent will AI be used and how.

ACKNOWLEDGEMENTS

This work was supported by the Silesian university in Opava under the student funding scheme, project SGS/10/2023 Media in Information Society.

The future is analogue?

Hynek Trojáněk

IURIDICUM REMEDIUM

iure@iure.org

KEYWORDS:

Digitalisation – Technology – Disability – Digital divide

ABSTRACT:

Digitalisation is usually described by journalists or politicians as something given and unchangeable – something that will simply happen. This inevitability is usually linked to a clear expectation: society will be freer, more informed, communication with the state will be easier. The time we save will then be invested in relationships or hobbies, in short, in more valuable activities than, for example, waiting for the post office or ‚going round the offices‘.

In recent years, however, the above ideas have shown major cracks. They have been illuminated in particular by the pandemic. On the one hand, it has intensified digitisation efforts, but at the same time it has revealed the negative effects that digitisation brings. Apart from the necessary collection of easily exploitable data, which is a by-product of all digitisation processes, it is in particular digital exclusion: a process in which whole groups of people literally drop out of society.

These are people who have difficult access to technology or who are unable or unwilling to work with it. This is not a marginal problem: „The data we have clearly shows that in all coun-

tries there are always some people who do not want to use digital technologies. This is roughly somewhere around 20% of the population. The exact proportion depends on the country, however, it's not a matter of some time and so the solution is not to wait until that generation dies out... It doesn't look like this is a path that will be at all meaningful in a few decades," Petr Lupac describes the digital exclusion in one of the available videos, adding that the digital divide is constantly renewing itself: most recently in artificial intelligence. „The first data we have now clearly show that the pattern of use of this technology is actually the same as it was with the internet or social media. It's mostly used by people for whom it's relevant at work. Then it's people with higher education, higher income, especially men," he says.

In addition to people who cannot use digital technology because of disability, for example, a significant proportion of the excluded are people who do not use technology for other reasons – the increasingly common desire for a digital detox or fear of data misuse.

Under the burden of data, with authors such as Carl Benedikt Frey reporting that while labour productivity is increasing, leisure time has only increased by ten per cent between 1900 and 2000, we will increasingly re-evaluate our coexistence with technology and the call for ‚analogue‘ will grow louder.

In Switzerland, the right to analogue has already been partially approved in one of the cantons in 2023. In Slovakia, a girl who was wrongly forced by the state to use digital technology in school was recently compensated. We also see the call for analogue in the ongoing debates about cash.

We are certainly not looking at a future without digital technologies – but the union with them will not be as fatal and as given as many would expect.

The Power of Human Communities vs Facebook Algorithms

Jiří Korčák

DEPARTMENT OF SYSTEM ANALYSIS, FACULTY OF INFORMATICS AND STATISTICS, PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

jiri.korcak@vse.cz

KEYWORDS:

Community – Algorithm – Facebook – Content distribution
– Curation – Engagement

ABSTRACT:

Our research delves into the intricate dynamics between the algorithms used by Facebook and the organic interactions within human communities. At its core, our study examines the algorithms governing Facebook's content delivery, focusing on the three principal ranking signals: the identity of the content creator, the type of content, and user engagement patterns with the posts. It sheds light on Facebook's commitment to Responsible AI (RAI) by discussing principles such as privacy, security, fairness, inclusion, robustness, and transparency. Also our research findings address the impact of Facebook's algorithms on users' political beliefs and behaviors. It discusses the extent to which these algorithms contribute to creating echo chambers, where users are predominantly exposed to opinions and content that reinforce their existing beliefs. This is contrasted with the potential effects of these algorithms on fostering diverse political discourse.

Additionally, the study explores the concept of collective intelligence within human communities. It argues that communal knowledge and interactions could serve as a potent counterforce to the algorithm-driven content curation on platforms like Facebook. This section emphasizes the value of human interaction and community-driven content in enriching social discourse and information dissemination. The study concludes by reflecting on the implications of these findings for the broader understanding of social media's role in shaping public opinion and facilitating community engagement. We suggests that while Facebook's algorithms play a significant role in content delivery, the power of human communities in shaping discourse and information flow cannot be underestimated. This study contributes to the ongoing discourse on the intersection of technology, social media, and human interaction.

Closing word

We would like to express our gratitude on behalf of the organizers to all speakers and participants of the Systems and Ethical Approaches – Responsible AI 2023 conference. Together, we have explored the transformative power of artificial intelligence, examining its integration into both our personal and professional lives. In an era marked by rapid advancements in AI and digital technologies, the insights from this year's conference have provided a diverse and comprehensive perspective on the implications of these technologies across various domains. There is a unanimous recognition of AI's transformative potential, coupled with an awareness of its inherent risks, emphasizing the need for responsible and ethical approaches in its deployment.

Delving into the ethical dimensions of AI development is crucial, highlighting the importance of considering external factors, data structures, and learning methods in AI applications. We must contemplate new approaches to responsible AI development in light of the increasing complexity of AI and its societal impacts. AI's transformative effects in both professional and personal realms are evident. While AI boosts productivity and efficiency by automating routine tasks, it also raises concerns about the future role of human workers and the potential erosion of essential skills. This paradigm shift calls for a balanced approach to technology integration, ensuring that AI complements rather than replaces human capabilities. For instance, AI's application in traffic systems can optimize the operation and maintenance of technological infrastructure. Predictive diagnostics illustrate its role in ensuring the reliability and se-

curity of critical transport systems, underscoring the need for responsible AI integration in critical infrastructure.

AI significantly impacts the information society, as evidenced by changes in users' information behavior, overload, retrieval, and the impact of AI on social media. The interplay between Facebook's algorithms and human interactions serves as a prime example. These algorithms can create echo chambers and influence political beliefs, contrasted with the potential of collective intelligence within human communities, also supported by social networks, to counterbalance algorithm-driven content curation. Additionally, the challenges of personalization, information verification, and decision-making should be addressed, highlighting the digital divide and the importance of equitable access to AI technologies. Related to this is the challenge of digital exclusion and the need for a more inclusive digital future. The digital divide calls for a reevaluation of our relationship with technology, advocating for a balanced, less deterministic integration of digital technologies in our lives.

In conclusion, the diverse perspectives presented at the conference converge to underscore the profound impact of AI and digital technologies across various sectors. From enhancing efficiency in telematics to ethical considerations in AI development, the influence of algorithms on social dynamics, and the challenges of information management in the digital age, it's clear that the integration of AI into various aspects of our lives requires careful consideration of ethical, social, and technical factors. The future lies in balancing the benefits of AI and digital technologies with the preservation of human autonomy, societal values, and equitable access to these transformative tools. This comprehensive approach will ensure that technological progress aligns with the broader goals of societal well-being and sustainable development.

Mgr. Ing. Tomáš Sigmund, Ph.D.



SILESIAN
UNIVERSITY
IN OPAVA



SKODA

WE WOULD LIKE TO THANK OUR PARTNERS