Alvolution - al and digital technologies in the European Union

Eva Poptcheva, Bram Vanderborght, Miguel Colom, Rūta Binkytė, Oana Balalau, Oana Goga, Hervé Debar, Marceau Coupechoux, Juan Herrera

AIvolution

AI and digital technologies in the European Union

renew europe.

AIvolution

AI and digital technologies in the European Union

Conclusions

Disclaimer

Considerations and recommendations expressed by the authors in "AIvolution" and/or in this white paper belong solely to their authors, based on their professional and scientific expertise, and are not necessarily endorsed by Renew Europe. Similarly, participation of the authors in the conference and/or in this white paper does not constitute endorsement to the political positions of Renew Europe.

Index

Eva Poptcheva MEP

Holds a PhD in Constitutional Law from the Autonomous University of Barcelona and she is a member of the European Parliament for the Ciudadanos delegation, within the Renew Europe parliamentary group. She is vice-chair of the Committee on Economic and Monetary Affairs.

Preface

Eva Poptcheva MEP

The potential impact of artificial intelligence (AI) on our economies is immense. According to some estimates global GDP may increase by up to 14% by 2023 as a result of the accelerating development and take-up of AI [1]. AI will drive this growth in three important ways: by increasing labour productivity, by making industrial processes more efficient and by creating a new virtual workforce capable of solving problems and self-learning [2].

This massive economic potential raises major questions. What role will the European Union play in this revolution? The internet revolution and the boom of internet technologies in the early 2000s took the EU wrong-footed. As a result, EU companies lag far behind the large US and China conglomerates in these fields. The same thing cannot happen with artificial intelligence.

On the other hand, the emergence of AI raises concerns on its potential disruptive effects on the economy. Some argue it could lead to the rise of massively scaled organisations and the overconcentration of wealth in some companies and sectors. The potential destruction of jobs is another cause of concern. Indeed, forecast by think-tank Bruegel indicates that 54% of jobs in the EU face probability of risk of computerisation within 20 years [3].

These concerns, in addition to the potential threats that abusive applications of AI could pose to fundamental rights and democracy, have led legislators in the EU to take action. The European Parliament and the Council of the European Union have recently found an agreement on the first ever legal framework on AI. With this landmark rulebook, the EU is becoming a pioneer in regulating AI, but is still lagging behind the US and China in AI innovation. Will the EU succeed in becoming a leader in AI rather than just a regulatory actor?

Shedding light on these matters and informing the public debate was precisely the objective of the event "AIVolution" which I had the pleasure to organise at the seat of the European Parliament in Brussels on the 16th of November of 2024. It gathered prestigious experts from different European universities and the private sector who are driving the AI revolution and studying it closely.

This document compiles some of the findings that the speakers presented during their interventions.

References

[1] PriceWaterhouse Cooper - The macroeconomic impact of AI https://www.pwc.co.uk/economic-services/assets/macroeconomic-impact-of-ai-technical-report-feb-18.pdf

[2] EPRS https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf

[3] The computerisation of European jobs - Bruegel https://www.bruegel.org/blog-post/computerisation-european-jobs

Section I
AI technologies

I.1 Robotics and AI

Bram Vanderbroght

A balanced approach to AI development is crucial for the EU's competitiveness and sovereignty. We should not only invest in AI algorithms, but also in the ecosystem in which it operates and the underlying (hardware) technologies it relies on as sustainable materials, advanced sensors, more energy efficient chips for advanced processing and communication technologies, novel application embodiments as robotics, wearables, IOT etc. This requires a multidisciplinary approach, not only focusing on the technological challenges, but also how medical, social and human sciences allow for a human-centered introduction in our society. The AI-driven applications will enhance various sectors such as healthcare, agriculture, and manufacturing. Moreover, they have as additional advantage that they have stronger intellectual property protection possibilities and are harder to replicate or relocate the production and manufacturing to other regions.

We should learn from our past mistake of outsourcing manufacturing to China and assuming that only the design is sufficient to keep in Europe. It is important to have the complete value chain available in Europe. Manufacturing is not only a source of economic growth and employment, but also a driver of innovation and quality, which are key for developing and deploying AI systems. By retaining and strengthening the value chain and its manufacturing base, the EU can ensure its technological sovereignty and autonomy in AI.

Europe starts from a strong position in robotics. In industrial robotics, it controls around one third of the world market, while in the smaller professional service robot market European manufacturers produce 63% of the non-military robots. Many robots do not use artificial intelligence algorithms for the control of their behaviour. Their control is dependent on known models, and, based on these models, calculations and input data, a complex algorithm defines the output and hence the behaviour of the robot. For example the impressive robots of Boston Dynamics do not use machine learning.

However, as such the system cannot learn. Artificial intelligence systems provide robotics application new impressive perspectives, but also come with complex challenges. For aspects like vision and speech recognition, deep learning has drastically outperformed traditional methods. These approaches typically require massive amounts of labelled data and computational power. Although data is typically abundant in robotics, labelling is sparse and expensive. For the control of robots, reinforcement learning often requires significantly more iterations than are feasible on real systems.

Therefore, a lot of the learning work is done in virtual environments, with challenges to transfer to the real world. Moreover, robots are interacting with the real world and challenges exist how to learn in a safe way, for the humans, for the environment and for the robot itself. Also explainability is important, so humans are able to understand and hence trust how the AI system came to its decisions.

References

[1] "ROBOTICS IN EUROPE — Why is Robotics important?" In EU Robotics. Available at: https://old.eu-robotics.net/sparc/sparc/robotics-in-europe/index.html

[2] "euROBIN Strategic Research Agenda (SRA 2024) — Executive Summary Version". In euROBIN Project. Available at: https://www.eurobin-project.eu/index.php/showroom/news/56-eurobin-strategic-research-agenda-sra-2024-executive-summary-version

[3] Brunke, Lukas, et al. "Safe learning in robotics: From learning-based control to safe reinforcement learning." Annual Review of Control, Robotics, and Autonomous Systems 5 (2022): 411-444.

[4] European Commission, Directorate-General for Research and Innovation, Vanderborght, B., Unlocking the potential of industrial human–robot collaboration – A vision on industrial collaborative robots for economy and society, Publications Office, 2020, https://data.europa.eu/doi/10.2777/568116

[5] Vanderborght, B. (2020). Unlocking the potential of industrial human–robot collaboration: A vision on industrial collaborative robots for economy and society.

[6] An Jacobs, Lynn Tytgat, Michel Maus, Romain Meeusen, and Bram Vanderborght (2019). "Homo Roboticus: Creating Synergies Among Humans, Technology, Science, and Art". IEEE Robotics & Automation Magazine.

I.2 Non-LLM key use cases of AI technologies

Miguel Colom

Artificial intelligence (AI) has regained a large popularity in the last years, certainly very deserved thanks to impressive achievements in robotics, biology, and very recently because of applications from generative large language models (LLM) such as ChatGPT, Bard, and others.

Generative LLMs are not only useful tools, but also they make us feel very curious about (and sometimes even afraid of) the underlying mechanisms they are based upon. At this point, probably the claim of Arthur C. Clarke about technology and magic is coming to the minds of many of the readers.

LLMs seem to deal with problems that were traditionally only the domain of human minds, from playing chess to understanding and writing text. This deserves a deep discussion about which characteristics are inherent to humans, but we will not go that far in this short text. Instead, we will glance at applications of AI other than LLMs that have a large impact in our daily lives.

We shall briefly comment on three areas with applications totally different from LLMs:

- *Biomedical*
- *Fundamental biology*
- *Fight against disinformation*

I.2.1 Biomedicine

In the biomedical field, machine learning (one of the branches of AI) has proven to be an excellent ally of humans, with applications in drug discovery, early diagnosis of diseases, and detection of tumors in medical imaging.

In 2020 there were many methods proposed to detect breast cancer, some of them with an accuracy which outperformed humans, such as the work of McKinney for breast cancer screening [1] or the one of Majumder [2].

As typical in machine learning methods, these artificial intelligence systems are trained by providing many positive and negative samples. The system is given images which show or do not show a tumor, and the model learns a function which hopefully is able to generalize to new cases and predict accurately.

Albeit the conceptual simplicity, training large systems such as deep neural networks is challenging, since the exact parameters needed to train properly the networks are not known in advance, and also because of the large volume of training data.

Much effort has been put in understanding large and deep neural networks, and the current research goes beyond the black box model, a system which is simply fed with many samples with the hope that it learns a function that indeed is able to generalize and predict, in some kind of brute force procedure which does not provide any insight on what is has actually learned.

On the contrary, sophisticated architectures are being proposed. Of many, we should cite the Transformers, which introduced in 2017 what is known as the attention mechanism [3], among other innovative techniques.

Transformers were designed for text analysis in mind (translation, mainly) and the attention mechanism allowed assigning weights to importance of each word relative to their position in the text for a better understanding of the sentences.

In 2020 the Vision Transformers (ViT) were introduced, and instead of operating in text, they are applied to images. In the biomedical field, they have been proved to have a very good performance to detect and localize tumors from medical images. We can cite the recent work of Feng [4] in 2023, which uses ViT to detect breast tumors with ultrasound images.

IA has also been proved very useful in early detection of autism in children [7] (and this is of major importance, given the benefits of quick intervention) and in the early detection of diseases such as Alzheimer's [8], just to cite a few applications.

I.2.2 Fundamental biology

Fundamental Biology has also benefited from AI, specifically in revealing the 3D structure of proteins, the essential bricks of life. They are chains of amino-acids that, after a process known as folding, get their characteristic 3D structure. Determining their 3D structure is absolutely required, since it is related to its function.

Biologists use techniques such as X-ray crystallography or nuclear magnetic resonance to obtain the 3D structure of proteins in the lab. Although this process provides the best confidence, it is very time consuming. Depending on the case, it could take ages to determine the structure of a complex protein!

The AlphaFold system developed by DeepMind in 2018 and later improved in 2020 helped in finding the structure of several complex proteins. In the version of 2018

AlphaFold took as input sequences of amino-acids that seemed to be correlated, assuming that they were close spatially. The different teams involved in this research defined a function whose input was a structure from the amino-acids and the output an energy related to the spatial structure. The lower the energy, the more likely the estimated structure corresponds to the real one. The goal of the system was therefore to find a 3D configuration of the proteins in order that the loss function was minimal, thus revealing the 3D structure of the proteins.

In 2020 AlphaFold was further improved adding, among other new techniques, the attention mechanism already introduced with Transformers [9]. And in 2021, the AlphaFold Protein Structure Database was launched, and since then it has revealed the structure of about 200 millions proteins. Without any doubt, this has been a major breakthrough in fundamental Biology, which will contribute to our understanding of several biological processes and also allow for synthesizing more efficient drugs.

I.2.3 Fight against disinformation

A totally different topic but yet with direct consequences on society is the fight against disinformation. The latest advances in AI have made it possible to forge new content easily. This has a positive side, such as helping artists in their creations or improving the writing and structuring of text, also for educational purposes.

However, as always there's a perverse side, which is taking advantage of these tools and techniques for disinformation purposes. Misleading information can be deliberately spread to manipulate society. For example, to influence election processes. The verification of the validity and origin of content is needed, and we researchers along with journalists and fact-checkers need to work together to debunk forged content.

Several initiatives and European projects contribute to the fight against disinformation by providing AI-based tools and methodology. One of them is VERA.AI (VERification Assisted by Artificial Intelligence), on which several international partners work together to develop methods and techniques for evidence retrieval, detection of deepfakes and other synthetic media, track forged content along social media, incorporate the feedback of experts to adapt AI models, and propose good practices. Most of these objectives involve AI.

Without going into the details, many deep-fake methods are based on generative adversarial networks (GANs), which are made of two competing neural networks. One tries to introduce the forgery, and the other tries to detect the forgery. During training, both networks compete until a good balance is reached: the resulting image or video looks natural and the forgery is not easily detectable. Several methods have been developed to detect deep-fakes, and we can cite, for example, the work of Rössler [10].

Making forged images and video has become a not-so-difficult task nowadays because of the availability and ergonomics of modern AI-based tools and methods. On the other hand, they leave traces that other AI-based detection methods can detect. This sort of cat-and-mouse race shows the good and evil sides of the discipline.

I.2.4 Conclusion

To conclude, it is clear that modern AI-based methods have a clear and positive impact in our society, from contributing to real medical and biological advances, fighting disinformation, and producing or helping arrive at new knowledge. The result is not only theoretical, but we benefit from applications for which we obtain an immediate benefit as society and individuals. All stakeholders need to be implicated, including citizens, scientists, regulators, and companies providing technology means in a responsible way.

As any other technology, improper or harmful use of AI-based techniques is certainly possible but, even so, the benefits of AI totally overcome the risk.

References

[1] McKinney, S.M., Sieniek, M., Godbole, V. et al. International evaluation of an AI system for breast cancer screening. Nature 577, 89–94 (2020). https://doi.org/10.1038/s41586-019-1799-6

[2] Majumder A, Sen D. Artificial intelligence in cancer diagnostics and therapy: current perspectives. Indian J Cancer. 2021 Oct-Dec;58(4):481-492. doi: 10.4103/ijc.IJC_399_20. PMID: 34975094.

[3] Vaswani, Ashish, et al. Attention is all you need. Advances in neural information processing systems, 2017, vol. 30.

[4] Feng H, Yang B, Wang J, Liu M, Yin L, Zheng W, Yin Z, Liu C. Identifying Malignant Breast Ultrasound Images Using ViT-Patch. Applied Sciences. 2023; 13(6):3489. https://doi.org/10.3390/app13063489

[7] Ortiz, Andrés, et al. Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. International journal of neural systems, 2016, vol. 26, no 07, p. 1650025.

[8] Chen, Junya, et al. Enhancing early autism prediction based on electronic records using clinical narratives. Journal of Biomedical Informatics, 2023, p. 104390.

[9] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[10] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, "FaceForensics++: Learning to Detect Manipulated Facial Images," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1-11, doi: 10.1109/ICCV.2019.00009.

I.3 Fairness in AI

Ruta Binkyte

AI fairness is a sociotechnical problem that needs an interdisciplinary approach. The idea is to have interdisciplinary in development, where social impact can be evaluated by social scientists, ethical approaches by philosophers, etc., and of course, domain experts also have an important role because fairness decisions are very domain-specific.

Fairness decisions often involve trade-offs and should be consulted by the stakeholders. A framework for identifying and involving stakeholders should be put in place. There is unavoidable arbitrariness to the fairness decisions, so people who will be directly affected by the system must understand and have a say in those decisions.

It typically requires the sensitive attribute to be present in the data set for bias evaluation and mitigation. Often sensitive attribute such as gender or race is not the data, because they cannot be collected or used due to privacy reasons. It could be reconsidered to measure and mitigate unfair biases in the data.

Section II
Data, disinformation and digital rights

II.1 Improving the quality of public debate

Oana Balalau

Nowadays, we are exposed to an abundance of data that we need to filter and interpret to make informed decisions. Researchers and developers have created exciting and powerful tools designed to assist journalists and citizens via AI and data management.

Journalists started using tools tailored for querying massive datasets to explore datasets containing leaks of millions of documents [1]. These datasets can be viewed as graphs, where nodes represent real-world entities, such as companies and people, and edges between the nodes are relations, such as person A working for company B. Journalists have found illegal or ethically questionable activities by following the money flow and connecting bank accounts to people. One of the tools used by the International Consortium of Investigative Journalists for querying and visualizing this data is Neo4j [2], a graph database. In contrast, more recently, other tools emerged, such as ConnectionStudio [3], which can ingest any semi-structured dataset (JSON, CSV, XML, etc.). However, much of the information available today is in the form of text (unstructured format), so natural language techniques (part of AI tools) are employed to extract mentions of entities and their relations [5]. These tools are used as lead finders to highlight potential connections that should be looked at in more detail. Many challenges remain, for example, entity disambiguation, the task of discovering

all the different names attributed to the same entity and all the different entities with the same name.

Fact-checking is another active area for journalists; well-established media will have a dedicated fact-checking team. Research has shown that fact-checking effectively reduces false beliefs [6]. However, the effects might only be long-lasting if the correct information is widely and repeatedly circulated [7]. Computer scientists have proposed a suite of solutions for fact-checking [8], varying from a completely automated solution that gives a true or false label to an affirmation to a partial solution that leaves the final verdict to a journalist [9], as the one currently deployed at RadioFrance. Completely automated solutions suppose there exists one evidence document, or a suite of evidence documents, which directly contain the truth value of the claim or from which we can deduce the truth value. Recent works have also promoted justifications, such as the list of supporting evidence or automatically generated paragraphs explaining why a specific label was given. However, these models can suffer from hallucinations; that is, they divert from the source evidence [8]. Automated fact-checking can become an asset for journalists; however, given the importance of the task, such tools should have a human in the loop that verifies if each step of the verification was done correctly and takes responsibility for the verification. In addition, the European Union should certify any such tool and inform the public of its limitations and strengths.

Finally, false information is not the only danger in today's public debates; fallacious arguments [10] are incorrect but compelling arguments that are often used as propaganda techniques [11]. Argumentation mining, the study of automatic extraction and reasoning over human arguments, has made significant progress in recent years [12] and has drawn significant industrial interest via projects such as IBM Project Debater [13]. Such arguments and techniques are more often used to convince the audience on more ambiguous or complex topics, such as ethical ones, for example, immigration and human rights. Both right and left leaning journals use fallacious arguments and propaganda techniques in political news reports, and citizens often have a strong engagement with such news [11]. While automatic methods can detect fallacies and propaganda more generally with some success [10], for citizens to accept the labels of such tools, they should have a good understanding of logic and argumentation. We note that AI can also generate propaganda; hence, it is widely speculated that we will be exposed to more such content in the future.

Based on the current trends in the use of AI by both positive and negative actors, we believe that education should be adapted to cover, at the appropriate levels in compulsory curricula across the EU, a basic understanding of AI workflows, data analysis, and interpretation, and finally critical thinking, in the form of logic and argumentation theory.

References

[1] https://en.wikipedia.org/wiki/Pandora_Papers

[2] https://neo4j.com/press-releases/2022-connected-data-fellowship/

[3] User-friendly exploration of highly heterogeneous data lakes. N. Barret, S. Ebel, T. Galizzi, I. Manolescu, M. Mohanty http://connectionstudio.inria.fr/

[4] https://en.wikipedia.org/wiki/Open_information_extraction

[5] Open Information Extraction with Entity Focused Constraints. EACL Findings 2023, P. Upadhyay, O. Balalau, and I. Manolescu

[6] The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom, Porter E, Wood TJ., Proc Natl Acad Sci USA. 2021

[7] The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain, and Canada, Carey, J.M., Guess, A.M., Loewen, P.J. et al., Nat Hum Behav 6, 236–243 (2022)

[8] A Survey on Automated Fact-Checking. TACL 2022, Z. Guo, M. Schlichtkrull, A. Vlachos

[9] Fact-checking Multidimensional Statistic Claims in French, O. Balalau et al., TTO 2022, https://team.inria.fr/cedar/projects/statcheck/?ref=medianes.org

[10] Breaking Down the Invisible Wall of Informal Fallacies in Online Discussions. ACL 2021, S. Sahai, O. Balalau, and R. Horincar

[11] From the Stage to the Audience: Propaganda on Reddit. EACL 2021, O. Balalau and R. Horincar

[12] Towards Argument Mining for Social Good: A Survey. E. M. Vecchi, N. Falk, I. Jundi, G. Lapesa, ACL 2021

[13] https://research.ibm.com/interactive/project-debater/

II.2 Disinformation, micro-targeting and political advertising in online platforms

Oana Goga

We all have been exposed to the fact that online platforms can create some real dangers in the real world. Online platforms have been blamed, for example, for the suicide of a 14 year old girl because of the content that she was seeing online [1]. It has also been said that online platforms led to the radicalization of Brazil [2]. And we know about Cambridge Analytica and the fact that they use personality-based ads to influence voters before the presidential elections of 2016 in the US [3]. So we can see that these online platforms can create some real troubles, both for human well-being and for our democracies.

There are three AI-related risks that contribute to the dangers of these online platforms.

- *First, we have AI-moderated content: we are using AI to detect hate speech or toxic content and all sorts of other things. And of course this AI can have this limitation, can have some biases.*

- *The second risk is related to AI-distributed content. We are using AI tools to figure out who would most likely engage the particular content and not with another.*

- *And this also brings other kinds of risk, as exposures to only certain kinds of content. And now there is a new risk which is of both AI-generated content and the fact that it is much easier to create content that might speak to certain groups of people.*

The focus of my latest research is on advertising technology. Advertising technology has all these three AI-related risks, and is problematic because it enables information targeting through online platforms. We have online platforms that are known for using AI algorithms to infer certain properties about users (what they are interested in, what is their behavior, etc.), and these platforms are exposing these users properties to advertisers that they can use to target with information.

Let's consider the example of an advertiser that wants to convince people not to get vaccinated. This advertiser can create one version of this ad that is designed for and targeted at people who are interested in alternative medicine; and the same

advertiser can create another version of this ad that is targeted at people who are from an opposing political party.

This technology is providing everyone with the ability to use users personal data to manipulate them. Because of this technology, risks go beyond disinformation, because truthful information or opinion can be also weaponized to influence public opinion.

Measurement methodology. One of the reasons by which many of these threats took us by surprise is that as external researchers, we do not have access to data, due to the closed nature of online platforms. We do not know what content (ads, posts) users are exposed online, and therefore we cannot assess the risks. Most of the existing literature in the space is only able to study things that are visible. For example, there are some small datasets on Twitter that we can look at and have an estimation of risk, but we are very limited in terms of data.

In order to overcome this challenge, we created a new measurement approach that is based on donations of personal data from citizens to science. We created a software tool that is able to observe the content people are seeing online and how they interact with them, and then send this data to our servers so we can actually see what posts, what ads they actually see.

We did two recruiting sessions with this tool: once in 2018 before the Brazilian presidential election, and once in 2020 in the US, again before the presidential election. What follows is a summary of the results we observed.

II.2.1 Is asking for transparency enough?

We know that these platforms have been misused in the context of election, so we want to have more transparency. One of the basic research questions, which was very popular initially, is, is transparency enough?

In the first study, we looked at the Facebook Ad Library. This was a central repository of political ads, where Facebook has promised to put all the political ads that are running on the platform. But we did not know what is their approach in putting political ads in this central repository. Our question was, is this Ad Library missing political ads? How many political ads are missed by the Ad Library?

Thanks to our data and the fact that we were actually able to observe on people's feeds whether they received political ads or not, we were able to see that half of the political ads we detected were actually missing from the ad library, at least in Brazil during the 2018 presidential election. The takeaway from this is that transparency is very important, but we do need to audit the transparency mechanisms that are provided by these platforms.

II.2.2 Can we apply restrictions on political advertising?

A second proposed solution was to apply restrictions on political ads. Our question was, can we do this in practice? For this, first we need to understand whether people actually agree on what ads are political and what ads are nonpolitical [5].

We took a dataset of 55,000 ads (ProPublica/Quartz) that were labeled by at least one user as being political. All these ads have at least three votes (from volunteers) on whether they are political or not. We looked on how many ads people agreed that they are political. In Fig. 1, the blue line shows the quantity of the ads where everyone agrees that they are political; the green line shows ads where people they disagree: some of them say that is political, some of them say that it is not.

Overall, in more than 50% of the cases people did not agree on whether an ad is political or not. So then, of course, the question was, what are these ads? Why people do not agree on whether these ads are political or not?

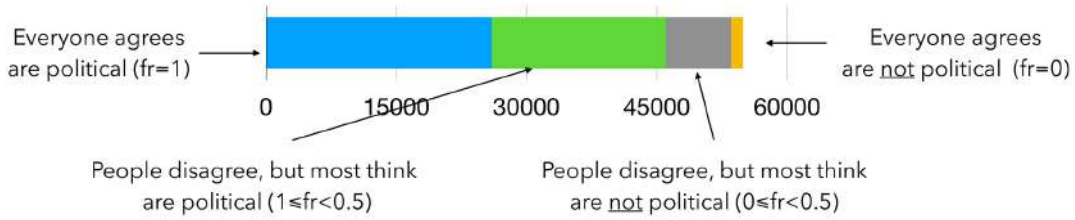

Fig. 1. Do people agree what ads are political?

We realize that there are ads from politicians, ads that talk about politicians — these do not create any disagreement. But then there are a lot of other ads that talk about social issues (e.g., immigration, civil rights, etc.), that do not mention a politician, and are not from a politician, but they discuss problems that are very important for a country. And if we look at these social issue ads, it is very hard to distinguish what is political and what is, for instance, humanitarian; what is activism, what is not activist.

Fig. 2. shows two ads, one from an activist NGO that is trying to promote action on immigration laws (left), and another ad that is trying to get money for homeless people (right). These sorts of ads are creating a lot of confusion.

If we are not agreeing on what is are political and nonpolitical, the moment that we want to apply a regulation on political ads, we will not know on which particular ads that we should apply regulation because we do not agree on on what is political.

II.2.3 What is the economic impact of banning micro-targeting?

Another question that we explored was whether banning micro-targeting could be a solution. For this, one important aspect to consider is what would be the economic impact of banning micro-targeting. We try to understand to which extent small and medium sized businesses are actually using micro-targeting on Facebook, based on our dataset. We linked this advertisers to their link to profiles, so we were able to infer what is the size of their company. As shown in Fig. 3, over 70% of advertisers on Facebook were small and medium sized businesses (SMEs), and they are also responsible for the majority of ad impressions; if we apply a regulation, they might suffer. It is thus important to assess the impact of regulation on them.

The next question to consider is, are businesses (SMEs and BEs) using micro-targeting?

Before going into this, it is worth to clarify a point about micro-targeting. Traditionally, when we are talking about micro-targeting, we always imagine that the platform assigns labels to its users: "this user is interested in coffee, this user is interested in sports"; they are grouping users according to some properties, and then allow advertisers to select the users they want to reach. This is advertiser-driven micro-targeting.

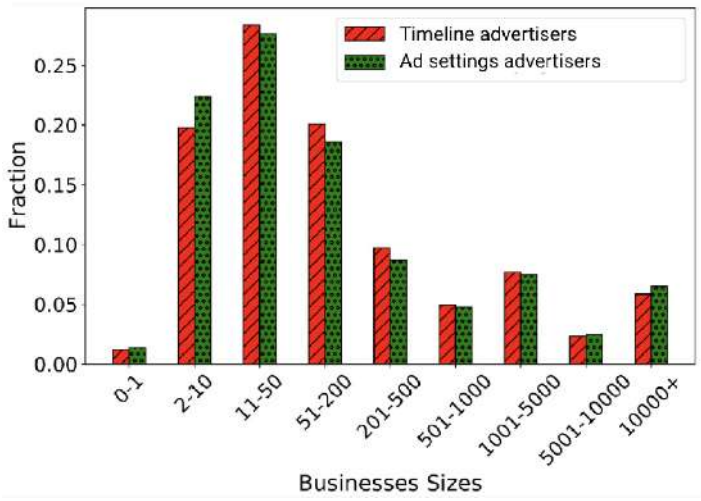

Fig. 3. To which extent small businesses advertise on Facebook? (SME: Small and Medium Enterprises; BE: Big Enterprises.)

This is one way of performing micro-targeting, but there is a different type of targeting which is algorithmic-driven micro-targeting, also called ad optimization. Here is not the advertiser that is listing the characteristics of the users to reach; it is the online platform that decides, based on the models and based on the ads, who are the users that are the most likely to be interested in this ad. Advertiser-driven targeting can be thought as explicit profiling of users, and algorithmic-driven micro-targeting is implicit profiling of users, because there is no need to put labels on users; it is just the AI who will estimate the users that are most likely to click.

Back to the question: do businesses use micro-targeting? The table at Fig. 4 shows the fraction of businesses using both micro-targeting variants: only 27% of small and medium sized businesses are using advertiser-driven micro-targeting.

	SME (%)	BE (%)
Advertiser-driven micro-targeting	27.7	30.5
Algorithmic-driven micro-targeting	72.3	69.5

Fig. 4. Do businesses use micro-targeting?

This was very surprising for us because in 2019, we had a similar study [4] that showed that businesses were using 79% of advertiser-driven micro-targeting, and 21% on algorithmic-driven micro-targeting. We observe now that they are relying more (72%) on algorithmic-driven micro-targeting: there is a substantial shift from advertisers

specifying their audiences, into advertisers outsourcing this task to the platforms, so that the platforms themselves are deciding who sees what ad. The takeaway from this is that we thought of banning micro-targeting in the traditional (advertiser-driven) sense, but there has not been a discussion of whether we can regulate this algorithmic-driven micro-targeting — which is replacing advertiser-driven targeting and becoming dominant. The question remains open on how to regulate this algorithmic-driven micro-targeting.

II.2.4 Is contextual advertising a safe alternative? Children targeting

Another idea was to explore whether contextual advertising is safe. What is contextual advertising? Whenever you are going on a sports site, you are going to receive an ad about shoes or sports clothes. This is not related to tracking or to profiling; it is just related to the content that you see.

Related to this, we try to see whether we can target children with ads. Our assumption is that we should not be able to, but maybe some features might allow us to do this.

We focus on YouTube because this is a platform that is most used by children. We realize that YouTube allows advertisers to place their ads on a particular video — and not just on groups of videos. Therefore, an advertiser can create a list of children-focused videos (which is very easy to do) and instruct the platform to only show its ads on these specific children-focused videos; this way children can be targeted through this technology with ads.

This placement-based advertising that may target children is a form of contextual advertising. But existing regulation on this matter does not seem conclusive. In particular, we observe that:

United States' COPPA does not prohibit advertising to children. Data collection (e.g., tracking) is restricted from children under 13 years old without verifiable parental consent; and online platforms' capabilities to serve profile-based ads to children is restricted. COPPA, however, does not restrict contextual-based advertising.

The EU's Digital Services Act forbids targeting children with ads based on profiling, but it does not forbid to target children through contextual advertising.

According to this, under current regulations, targeting children through placement-based advertising is legal, both in the EU (DSA) and in the US (COPPA). It is not clear whether legality of this practice is deliberate, or it comes from an insufficient understanding, by lawmakers, of the implications of placement-based advertising on children.

More in general, placement-based advertising can also be used in other contexts — not only children targeting—, such as health, misinformation, etc. Placement-based

advertising allows advertisers to choose whatever content they want on YouTube, and specifically target users that watch some particular videos. This placement-based micro-targeting could be thus even more dangerous than what we thought about profiling-based micro-targeting.

II.2.5 What can we learn from malicious advertisers?

One last thing that we explored is what are malicious advertisers telling us about how they behave. We looked at the ads that were seen by the Internet Research Agency before the 2016 presidential election; Fig. 5 displays the main characteristics of these ads.

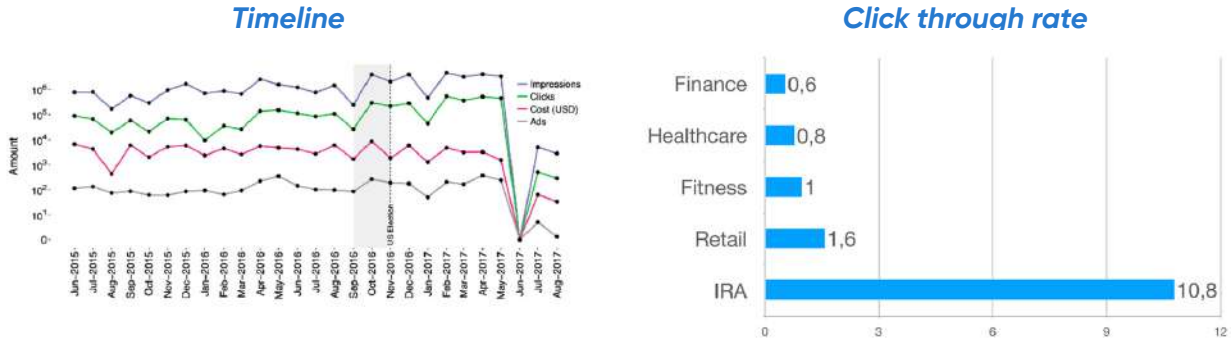

Fig. 5. The Internet Research Agency ad campaigns.

There are two relevant observations on Fig. 5.

The first one (on the left) is that their campaigns run over two years, which could be relevant for defining political adds. It is not just before elections that organizations like IRA try to influence people.

The second one is that their ads have a click-through-rate (CTR) 10x to 20x higher than normal ads (on the right). The click-through-rate metric indicates the number of clicks, over the total number of users that viewed the particular ad; it shows the effectiveness of the ad. Such a high value indicates that IRA ads were highly effective.

The CTR metric, as proxy of effectiveness, provides a very interesting information that could be exploited to detect malicious campaigns. In terms of transparency, if you have an ad library with all the ads, CTR would be a useful information to detect the ads that would be more likely or that we should check because they have a higher effect on people.

II.2.6 Access to data is key for measuring and mitigating risks

Last point to make is that access to data is key for measuring and mitigating risks. Until now, platforms have provided some APIs. However, none of the research presented in this contribution relies or used the APIs provided by these online platforms.

In this context, the Digital Services Act, and the delegated act that is allowing vetted researchers to ask for more private data from platforms, are one of the most important tools for researchers at this point, and maybe other NGOs as well.

Platform-provided data and alternate data access. We have been using alternate data sources; and they can be either obtained through what is called "scrapping" (i.e., a tool connects to a Web page, and tries to extract the content on the Web page), or we ask users to donate their personal data to science. But these techniques are against the terms of service of platforms; in the US, researchers have been pursued in justice because they created these automated "scrapping" tools.

These alternate data sources are very important for researchers, even if we have the Delegated act, even if we have access from platforms to platform-provided data; because these are the only ways that we can actually audit what platforms are sending us. Researchers do need solid legal frameworks to protect them so that they can perform independent audits of online platforms.

References

[1] "British Ruling Pins Blame on Social Media for Teenager's Suicide". The New York Times, October 1, 2022. https://www.nytimes.com/2022/10/01/business/instagram-suicide-ruling britain.html#:~:text=New%20York%20Times-,British%20Ruling%20Pins%20Blame%20on%20Social%20Media%20for%20Teenager%27s%20Suicide,a%20more%20than%20minimal%20way."

[2] "How YouTube Radicalized Brazil". The New York Times, August 19, 2019. https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html

[3] "50 million Facebook profiles harvested for Cambridge Analytica in major data breach". The Guardian, March 17, 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[4] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau, A. Mislove (2019): Measuring the Facebook Advertising Ecosystem. Proc. NDSS'2019. https://www.ndss-symposium.org/ndss-paper/measuring-the-facebook-advertising-ecosystem/

[5] V. Sosnovik, O. Goga (2021): Understanding the Complexity of Detecting Political Ads. Proc. Web Conference 2021 (WWW'21). https://arxiv.org/abs/2103.00822

Section III
Digital infrastructures

III.1 AI and cybersecurity

Hervé Debar

The emergence of Artificial Intelligence is creating new challenges in cybersecurity, as well as offering new opportunities [1]. The recent emergence of Large Language Models (LLMs) and their increased use in many personal and professional topics (generating documents, answering complex questions, generating code, etc.) is significantly changing the attack surface and offering new opportunities for attackers and defenders [2].

At this point in time, the question is not whether AI systems will be attacked, but how, and what the impact of these attacks will be.

III.1.1 Bias and hallucinations – increasing the attack surface

One of the well-known and studied issues related to artificial intelligence is bias. Depending on the learning mechanisms, and particularly the data used for training, AI algorithms may exhibit bias. This is for example well-known in image recognition tasks or in social network recommendation algorithms. Curating training data is particularly tedious and difficult, and is also hard to automate. For example, training an LLM to

generate code is very likely to include vulnerable or unsafe code in the training dataset, vulnerable code that will be considered as valid as safe code.

The difficulty in this aspect is that on one hand, AI fails at tasks that would be easily performed by humans. On the other hand, it also generates results that are easily accepted by humans. In the case of LLMs, this is known as hallucinations, the capability to invent false information, that looks credible, such as false citations or journal articles.

One of the key issues is whether these bias and hallucinations can be controlled by attackers [3], inserting some form of backdoor that could be triggered at will (by specific requests, for example). One particular risk is when using LLMs for code generation, where such code could be constructed to include backdoors or vulnerabilities [4][5]. Studies of adversarial machine learning address these issues and aim to provide solutions for model integrity and safety.

It is thus likely that including AI in existing systems, and developing new AI algorithms of significant complexity, will increase the attack surface of digital infrastructures.

III.1.2 Using AI for attacks and defense in cybersecurity

Artificial Intelligence is providing opportunities for attackers and defenders alike.

It enables automation and speed on both sides, in attack and defense, in exploring as many vulnerable paths as possible and in blocking automatically such exploration. While automation enables faster attacks, it also limits the attacker to vulnerabilities that can be automatically exploited. Unfortunately, there are enough of these that engineering mechanical attacks will have a significant impact on digital infrastructures.

Network and infrastructure deployment automation, as currently deployed in 5G networks and already available in cloud environments, will facilitate "on demand computing and networking". It will also require that these deployments include automated protection and defense against cyber-attacks; this may require new regulations beyond NIS and NIS2.

One of the areas where AI has already been used and is likely to improve the situation is in attack detection [6]. Malicious behaviour detection is an important research topic, that should lead to advanced security sensors for network and system protection, detection and response. Further work leveraging AI for large scale anomaly detection could be devoted to detecting and shutting down malicious digital infrastructures, currently deployed and operated for profit by criminal groups.

Similarly, AI is likely to enable the development of new risk assessment methods, applicable to complex interconnected environments. This will support better protection measures, as well as insurance against cyber-threats and regulatory compliance.

III.1.3 Understanding AI

One of the hot research topics today is explainable AI, or how to explain the results produced by artificial intelligence algorithms. This is seen as one of the keys towards the ability to certify artificial-intelligence-based digital systems [7].

There are in fact two aspects to explainable AI. The first one is "local explanation", or the ability to explain why an AI model gave a specific answer to a specific stimulus (answering the question "why is this specific decision reached now"?). The second one is "global explanation", or the ability to explain globally what the model has learned (or not) from the input data and training process.

Unfortunately, current explanation approaches might also increase the attack surface and are probably as insecure as AI algorithms. For example, different explanation methods might produce different results. We also cannot guarantee that explanation methods actually explain the model, and not contextual information.

Successfully addressing both aspects is one of the keys for ensuring that AI-enabled digital systems meet the needs and values of the EU. Significant effort must be deployed to ensure that production-grade models are indeed demonstrated compliant with the existing and future regulations. As we develop these regulations, such as the AI Act, we must at the same time develop the tools that will enable enforcement of these regulations.

Understanding AI models and algorithms is also a way towards certification, as it requires that we are able to assess what AI models do, and what they do not do, with a high degree of likelihood.

III.1.4 Economic aspects

A final issue with current AI is that creating and using models might be within the reach of only very few organizations, that have the ability to collect, curate and store the data

needed, to develop and run the required algorithms (requiring very expensive hardware and a lot of energy to power it), and to provide access to the general public. For example, it is very likely that only the GAFAM and the BATX have the capability to buy enough storage and Nvidia processors to store the data and run training programs for several weeks, possibly months.

This issue is more related to economics and sovereignty, and there is a clear need for Europe to host data and tools for AI that will be accessible to EU citizens and companies. There is also a notion of sustainability here, as there is a significant shortage of processors able to run these learning algorithms. It remains to be seen

whether actions such as the Chip Act will provide the EU with access to sufficient computing power.

Not having access to AI models and algorithms creates a new kind of economic dependence on these tools. We do not have an EU-based Large Language Model, for example. And the public competition for this seems to focus in the US, although it is likely that other superpowers are developing the same technologies. Beyond economics, we also need skilled people. There is a significant need for trained personnel, both for AI and for cybersecurity. Skills development is one of the keys for maintaining a significant attractiveness of both fields in the EU.

Another aspect is AI licensing. Comparable to the open-source software movement, there exist Open Source IA initiatives and announcements. There are however significant hurdles to this, as already mentioned. The cost of storage and computing is a significant barrier that open sourcing will have difficulties challenging. A second barrier is the brain power required to tackle the complexity of these models, and to clean the training data. Open source software is defining our digital infrastructures today. We should ensure, through technology and regulation, that open and trustworthy AI defines them tomorrow.

References

[1] Apostol Vassilev (NIST), Alina Oprea (Northeastern University), Alie Fordyce (Robust Intelligence), Hyrum Anderson (Robust Intelligence). Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations. NIST AI 100-2 E2023. January 2024. https://doi.org/10.6028/NIST.AI.100-2e2023

[2] Cinà, A. E., Grosse, K., Demontis, A., Vascon, S., Zellinger, W., Moser, B. A., ... & Roli, F. (2023). Wild patterns reloaded: A survey of machine learning security against training data poisoning. ACM Computing Surveys, 55(13s), 1-39

[3] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. arXiv preprint arXiv:2311.05232.

[4] Ken Thompson. Reflections on trusting trust. Communications of the ACM, 27(8):761–763, 1984.

[5] He, J., & Vechev, M. (2023, November). Large language models for code: Security hardening and adversarial testing. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (pp. 1865-1879).

[6] P. Mishra, V. Varadharajan, U. Tupakula and E. S. Pilli, "A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection," in IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 686-728, Firstquarter 2019, doi: 10.1109/COMST.2018.2847722.

[7] Minh, D., Wang, H.X., Li, Y.F. et al. Explainable artificial intelligence: a comprehensive review. Artif Intell Rev 55, 3503–3568 (2022). https://doi.org/10.1007/s10462-021-10088-y

III.2 6G: Stop or again?

Marceau Coupechoux

Current visions of future communication networks draw a utopian (or dystopian ?) world made of immersive experiences, an "Internet of Senses" based on haptic communication, digital twins, artificial intelligence, etc. Is all of this really virtual? No, all of this is very real; it's made of cables, antennas, power amplifiers, processors, batteries, air conditioning, concrete, and tons of electronic waste. First message: ICT is not virtual. And all of this has an impact on our environment.

What do we know today about the environmental impact of digital technology? It is estimated that in 2020, digital represented between 2 and 4% (Freitag et al. 2021) of global greenhouse gas emissions. ICT emissions increased at a higher rate than global emissions (perhaps with a factor of 2 between 2002-2012). The overall picture is relatively well known for greenhouse gas emissions or energy consumption, much less for other impacts, such as water use or the depletion of non-renewable natural resources.

Can we outline perspectives for the next 10-20 years ? It is difficult to conceive possible trajectories because of unknown new usages, new technologies and uncertainties about historical trends. In the literature, there is a relative consensus that data traffic and the number of connected objects should continue to grow if nothing is done. The energy efficiency gains and the increasing use of renewable energy are far from enough to achieve the +1.5° target. ICT emissions must indeed reduce by 42% in 2030, 72% in 2040 compared to 2010, and be net zero in 2050. If nothing is done, digital will represent 35% of emissions in 2050. Second message: ICT is part of the problem.

Did 5G deliver on its promises? It's hard to answer because we lack open data. 5G is between 4 and 10 times more energy efficient than 4G, mobile data traffic has however doubled in 2 years. So it is clear that we are not heading towards a drastic reduction in energy consumption. You might say: yes, but we are using more renewable energy. But no energy is completely clean or infinite. Their production is also notoriously insufficient. We will thus have to question certain priorities: Do we still want to use scanners for our health or shop in the metaverse ? Third message: We need to assess the impact of 5G, and for that, we need public data.

Is there a positive effect of ICT on other sectors? The digital industry has pushed the notion of enablement effect, ICT for Green, smart X, meaning it has tried to show that ICT could have positive indirect effects in different sectors of the economy through substitution. In fact, there is no scientific evidence that, at a global level, digital is beneficial in terms of energy consumption. A study by Lange et al. in 2020 even suggests that digitization has been associated with greater energy consumption so far. Fourth message: until proven otherwise, there is no win-win scenario; in other words, digital transition contradicts ecological transition.

Is 6G on the right track? 6G is currently being defined. However, it seems that the visions are not up to the environmental challenge. Most research programs are full of "sustainability" but never question a model that has seen a continuous increase in the amounts of exchanged data and the number of connected objects. The strategies of different actors are never questioned. The reason may be that the definition of 6G was entrusted only to market players when the civil society should have been massively involved. Fifth message: 6G is not on the right track from an environmental point of view.

As a conclusion, mass digitization for the past 25 years has not reduced emissions; anyone can see that. Energy efficiency improvements have not so far addressed the environmental crisis, and there is no reason they will in the future. The reason is probably that they often come with a rebound effect that nullifies the gains. Another reason may be that they do not question our growing consumption of data and connected objects. Sobriety, on the contrary, aims to achieve collective reduction goals (cf. F. Flipo, AOC 2023). It is a more interesting concept because it leads us to question our lifestyles, collectively ask what trajectory we want to follow, what our essential needs are, and design communication networks accordingly.

References

[1] Charlotte Freitag, Mike Berners-Lee, Kelly Widdicks et al. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns, 2021, vol. 2, no 9.

[2] Steffen Lange, Johanna Pohl, Tilman Santarius, Digitalization and energy consumption. Does ICT reduce energy demand?, Ecological Economics, Volume 176, 2020.

[3] Fabrice Flipo, Efficacité, effet rebond, sobriété, AOC, Nov. 2023.

III.3 The global competition for mineral raw materials in the development of AI capacities

Juan Herrera

Artificial intelligence (AI) heavily relies on a variety of mineral raw materials that are essential for manufacturing a wide range of goods and services in everyday life, all increasingly powered by AI systems. The demand for increasingly sophisticated equipment leads to higher consumption of critical raw materials. As global demand continues to rise, it intensifies the competition among nations and companies for secure and reliable access to mineral reserves and supplies. There is a big gap between where we are now (decapitalisation of fossil fuels) and the next generation of new possible solutions (which are not yet here) in the next 10 years.

The world we live in is changing. Population growth, decarbonisation, digitalisation, diversification and the implementation of circularisation, are altering the minerals supply, redefining mining models and creating new opportunities. This means that mining in the future will be widely different from today.

Population growth: World's population is predicted to rise from 7.8 billion in 2020 to 9.9 billion by 2050, more than 25% since 2020 and quadrupling the population in 1950 (2.5 billion). 90 % of those will live in developing countries: China and Asia-Pacific, India's subcontinent, Southern Africa and South America.

The demographic group born between 1994 and 2010 represents 25.9% of the world population and are true 'digital natives' who conceive work and consumption differently from their predecessor generations.

In addition to the global growth, there is a constant population displacement to mega-cities, a rise in living standards and much bigger investment needs in infrastructure (railways, roads, transport systems, power generation facilities, water and sanitation networks, and others) intelligently managed. All this has brought an accelerated social change and growth of development poles and a multiplication of innovation capacities and economic opportunities, implicating a completely different consumption of raw materials from these huge population centers in constant growth.

Decarbonisation: Since the adoption of the Paris Agreement at the COP21 in 2015, 70 % of the world's GDP has gone to decarbonized economies. In 2020, the EU established by 2030 the mandatory reduction of national emissions to <55% compared

to 1990 levels and guaranteed zero emissions in 2050. Also in November 2020, Chinese President Xi Jinping declared China's commitment to reaching net zero emissions by 2060 at the UN and just a few months later, in April 2021, US President Joe Biden rejoined the United States into the Paris Agreement and announced a commitment to reduce emissions by 2030 between 50% and 52% from 2005 levels, and the complete decarbonisation of the American economy by 2050. Since then, the three largest economies have stepped on the accelerator to a more sustainable world with millionaire plans, taking advantage of their economic recovery programs after COVID-19. According to calculations by Bank of America, the economic impact of climate change could reach 69 trillion dollars (58.4 trillion euros) by the end of this century, and the investment in the energy transition must increase to four trillion a year to reach the objective of keeping global warming well below 2 ° C and continuing efforts to limit it to 1.5 ° C.

Digitalisation: Digital technologies are the backbone of a net-zero emissions future, and any lower-carbon pathway will increase the overall demand for minerals. The fifth industrial revolution transforms companies into intelligent organizations to achieve new production models with full interaction between humans and machines, improving productivity and efficiency in the production processes of the industry. The development of the digital and decarbonized economy will require in the next 30 years the supply of more raw materials than Humanity has extracted in 70,000. For mining companies, the greatest challenge of the XXIst century will be satisfying this enormous growth in worldwide demand.

Diversification: In the context of the current global development, the demand for minerals will increase by nearly 500% by 2040. Clean energy technologies require more minerals than fossil fuel-based alternatives. The types of mineral resources used vary by technology, but developments include more and more different materials with different properties, which means more and more different minerals in the composition. As countries transition to clean energy, the demand for minerals, and particularly the demand for critical minerals and metals, will continue to grow. But together with the clean energy transition, the supply chain vulnerabilities (particularly after the COVID-19 pandemic), the geopolitical tensions, the technological innovation (like the development of 5G and quantum computing), the ongoing digital transformation, and the rapid growth in electric vehicle production, are fueling this rising diversified demand. Innovations in materials science, such as finding substitutes creating synthetic materials, or improving recycling and recovery rates of critical minerals can help mitigate or at least reduce the demand for critical minerals. Additionally, advancing technologies that require fewer or no critical minerals are other avenues to explore in the coming years.

Circularisation: A circular economy aims to reduce waste, prolong the life of materials at the highest possible value, design products for materials to be cycled back into the economy and regenerate nature. It is much more than recycling. In traditional linear business models, critical minerals are extracted, then used to manufacture a product and ultimately discarded – recovery or recycling of critical minerals is very limited. However, such models are no longer compatible with current supply chains, growing demand, and existing and future sustainability objectives and regulatory

requirements. The volume of end-of-life green infrastructure assets and electronic waste containing critical minerals is expected to rise sharply in the coming decades. According to the IEA, recycling will reduce the need for primary minerals by about 10% by 2040. In other words, increased production of minerals and metals will be required over a long period of time.The AI revolution will be linked to the transition to a net zero and digital society. But at the same time, they will continue to accelerate and increase our reliance on minerals, particularly on the critical ones. The five vectors described are reshaping the future of the raw materials industry as the fundamental provider. The development of a circular economy will reduce this reliance, improve the resilience and sustainability of organisations, as well as offer new commercial opportunities. AI will help companies change their approach to the manufacture and use of products and adopt new business models and digital technologies, but the need for more raw materials, especially the critical ones, will remain increasingly high.


renew europe.