



ARTIFICIAL

INTELLIGENCE

SOCIAL EUROPE DOSSIER

ARTIFICIAL INTELLIGENCE

SOCIAL EUROPE DOSSIER



Social Europe Publishing

Berlin, 2020

Copyright © 2020

This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DIII111 (‘Deutsches Internet-Institut’) and the Friedrich Ebert Stiftung.

All rights reserved.

No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the author, except for the use of brief quotations in a book review.

**FRIEDRICH
EBERT** 
STIFTUNG

weizenbaum
institut

ARTIFICIAL INTELLIGENCE, WORK AND SOCIETY

Artificial intelligence is permeating a wide range of areas and it is bound to transform work and society. This dossier, published in cooperation with our partner Friedrich-Ebert Stiftung and Weizenbaum Institute, addresses possibilities and challenges of AI. Above all, it asks what needs to be done politically in order to shape this transformation for the sake of the common good.

AI and work

AI has conjured up a dystopia of robots displacing human workers from employment. Some have predicted very large-scale job substitution but others question whether such a predetermined outcome can be envisaged: whether jobs are lost and how they are changed depends on whether workers are involved in the decisions that are made. Similar concerns apply to issues of recruitment and monitoring of workers: will AI data serve a ‘surveillance capitalism’ or could it assist workers in the performance of their jobs if they have more power to influence the outcome?

AI and society

AI raises wider questions about the society in which we live and that of the future. Market-research institutes foresee huge efficiency gains, but are these credible and, if so, how will such gains be distributed? Feminists and anti-racists have expressed concern that the algorithms on which AI depends unconsciously embed the social prejudices of their human authors. Issues of privacy and civil liberty surround the possession and control of the data mined by AI. How education must change so that citizens can feel empowered rather than alienated by AI is also at stake—as is the ever-present issue of where AI fits in meeting the existential challenge of climate change and biodiversity loss.

ONE

WHEN MACHINES THINK FOR US: CONSEQUENCES FOR WORK AND PLACE

JUDITH CLIFTON, AMY GLASMEIER AND MIA GRAY

Will artificial intelligence affect how and where we work? To what extent is AI already fundamentally reshaping our relationship to work? Over the last decade, there has been a boom in academic papers, consultancy reports and news articles about these possible effects of AI—creating both utopian and dystopian visions of the future workplace. Despite this proliferation, AI remains an enigma, a newly emerging technology, and its rate of adoption and implications for the structure of work are still only beginning to be understood.

Many studies have tried to answer the question whether AI and automation will create mass unemployment. Depending on the methodologies, approach and countries covered, the answers are wildly different. The Oxford University scholars Frey and Osborne predict that up to 47 per cent of US jobs will be at ‘high risk’ of computerisation by the early 2030s, while a study for the Organisation for Economic Co-operation and Development by Arntz *et al* asserts that this is too pessimistic, finding only 9 per cent of jobs across the OECD to be automatable.

In a new paper, we argue that the impact of AI on work is not deterministic: it will depend on a range of issues, including place, educational levels, gender and, perhaps most importantly, government policy and firm strategy.

Highly uneven

First, we challenge the commonly held assumption that the effects of AI on work will be homogeneous across a country. Indeed, a growing number of studies argue that the consequences for employment will be highly uneven. Place matters because of the importance of regional sectoral patterns: industrial processes and services are concentrated and delivered in particular areas. At present AI appears to coinhabit locations of pre-existing regional industry agglomerations.

Moreover, despite globalisation, national and local industrial cultures and working practices often vary by place. Different cultural work practices mean that, once deployed, the same technology may operate distinctly in diverse environments.

Secondly, education matters. Generally, jobs occupied by less-educated workers are more susceptible to the impacts of AI and automation, compared with better-educated peers performing more complex and discretionary tasks. For example, in the financial and insurance sectors repetitive, data-intensive operations may be more automatable in the US than in the UK, due to the differences in average education levels within these professions. Another example is legal services, where those in paralegal, less-skilled occupations are at most risk of displacement.

Thirdly, it appears men's jobs are currently more vulnerable to automation—especially those requiring lower educational attain-

ment, since these tend to be routine industrial tasks amenable to mechanisation. This may however change in the future.

Women dominate many care jobs in ‘high touch’ occupations, where emotional and cognitive labour are significant. These jobs appear more resistant to technological encroachment, as they involve face-to-face work. In the medium term, though, emerging applications aim to augment even these service functions with machine assistance and are likely to interact with and produce new gendered divisions of labour.

Narrow focus

Fourthly, the consequences of AI on work will depend, crucially, on policy and the firm. Acemoglu and Restrepo argue that productivity increases could outweigh the displacement effect of technologies under the ‘right’ type of AI: if governments actively support AI which enhances jobs, rather than AI which seeks to eliminate jobs, the outcome could be positive overall.

To do this well, government also needs to accompany AI with social policy. Governments have started publishing AI policies in the last few years. But a comparative analysis of government AI strategies shows that, to date, the great bulk of policy has focused narrowly on economic gains, with very little attention paid to social issues. Yet understanding the latter is a precondition of societies being able to evaluate, and regulate, new applications of AI.

Firms, too, can opt to promote the ‘right’ type of AI—or not. Meanwhile, they may increasingly turn to AI to support recruitment.

This could be problematic, since AI algorithms have been found to contain embedded gender and racial biases. The use of such tech-

nologies as facial and voice recognition, automated screening of *curricula vitae* and targeted profiling may inadvertently reduce the pool of eligible job-seeking applicants in profoundly prejudicial ways. If businesses put these to use for recruitment purposes, the distribution of job opportunities could be profoundly affected, and AI might reproduce pre-existing biases around gender, ethnicity, and class.

Two paths

At its starkest, we see two paths forward. Fuelled by scare tactics and the ‘great unknown’, consulting firms are pushing companies to jump on the AI bandwagon, to avoid becoming economic ‘lag-gards’. Each consultancy is carving out a niche toward distinct trajectories, from relying on cutting costs to eliminating low-skilled labour—and encouraging government AI policies to focus on economic gains.

Another path is however possible. The potential exists for AI applications to enable the reskilling of existing workforces, thus allowing workers to use their skills alongside new technologies. AI and associated technologies can be used to help transform education and health and, even, attain peace.

There is nothing preordained about how AI will be deployed. The application consequences of these technologies will reflect choices made at the organizational, political and societal levels. The future of AI is too important to be left to technology specialists. Social scientists, lawyers of technology and experts in the ethics of technology need actively to engage in shaping and structuring its development and adoption.

This article is based on a collection of articles on AI and work in the Cambridge Journal of Regions, Economy, and Society, volume 13, Issue 1, 2020.

Judith Clifton is a professor in the Faculty of Economics and Business Science, University of Cantabria (Spain), and a visiting scholar at St Antony's College, Oxford.

Amy Glasmeier is professor of economic geography and regional planning at the Massachusetts Institute of Technology.

Mia Gray is a senior lecturer and fellow of Girton College, Cambridge.

TWO

ROBOTS WON'T MAKE US REDUNDANT

LARS KLINGBEIL AND HENNING MEYER

Social democracy emerged from the labour movement in the 19th century. Work has always been the focal point of social-democratic politics. In recent years, however, the role of work has become discussed increasingly narrowly and defensively. Whether in the debates about digitalisation or earlier on globalisation, work has always appeared under pressure. We should take this discussion in a different direction.

We think this mantra is false. Although globalisation and digitalisation present us of course with new challenges, the significance of work in society is not diminished. On the contrary. If we are right in shaping the change that lies before us, the work of the future becomes one of the most effective instruments of social policy.

In the 1990s and 2000s, the dominant discourse, in Germany for instance, was that relocation of production and global competition would jeopardise jobs and wages. In the recent debates on digitalisation, some observers have even anticipated a labour-market apocalypse. The fear is that robots and artificial intelligence could make human labour almost completely redundant.

New opportunities

Forecasts of how many jobs will be lost in the future vary widely. The honest answer is that no one knows exactly how digitalisation will work out. What all experts agree on, however, is that the work of the future will lead away from routine and towards more creativity. In consequence, through this shift the socially-transformative potential of work grows rather than diminishing. This opens up new opportunities.

These days in Germany, industrial policy is finally being argued over again. This discussion is long overdue. The role of the state in the economy was for a long time interpreted too defensively. It must not be the role of the state merely to correct market failures. Rather, it is a question of creating markets themselves and shaping the economic process politically. Our society should not be subordinated to the economy; rather, the economy should adapt to the ideals of our society.

From an offensive industrial policy, good jobs, new technologies and social prosperity result—in that order. Those who want to solidify opposition to climate and labour-market policies and stick to the *status quo* will end up losing the most. Moreover, without adherence to the value of labour, a modern industrial policy is inconceivable. Finally, it is only through skilled jobs that new technologies are created to address the major problems of our time.

This also applies to the area of digitalisation. Data policy and the development of artificial intelligence will be decisive for jobs and growth. The global race has long since opened. For us, it cannot be a question of whether but only of how. The state must put itself in the driving seat and aggressively push for the onset of artificial intelligence in the economy and science and also in politics.

Ageing society

In the services sector, too, we need an offensive concept of work and a political strategy. From childcare to social care, our public services need to be upgraded through more and better work. An ageing society cannot allow itself in the long term a weakening welfare state or an education system worthy of improvement.

The renewal of the welfare state or the improvement of the education system is not achievable without more and better work. How can the shortage of childcare places be eliminated without more motivated educators? How can we become more attentive to individual needs in schools without more teachers? There is only one answer to this: it cannot be done without more and better work.

The renovation of the social arena benefits from more staff with better social competence. The education system should put the creative and problem-oriented skills of the future more strongly into focus.

Good work will therefore continue to be the foundation of our prosperity and an important indicator of the quality of our life together. If we continue to esteem work and shape it in a well-aimed way, we can make our society a better one. A society in which cohesion and togetherness have a firm place and a new prosperity opens up.

It is therefore time to address more proactively in the public discourse the significance of work in shaping our social future. It is the basis for mastering the great challenges of our time and at the same time a competitive advantage if the change at hand in the world of work is framed correctly.

Lars Klingbeil is since 2017 secretary general of the Social Democratic Party of Germany. He has been a member of the Bundestag for Rotenburg I—Heidekreis since 2009.

Henning Meyer is a social scientist and member of the SPD's Basic Values Commission.

THREE
INTO A NEW ERA OF WORK

DANIELA KOLBE

In the automation and digitalisation we experienced hitherto, people were given a machine—such as a laptop with the usual office software, a 3D printer or a computer-numeric-controlled milling machine—which they could use to perform their job. Knowledge and communication became more mobile. At the same time, the new machines made possible customised production.

Artificial intelligence (AI) systems enable machines to work with people. As they are being introduced into the workplace, new kinds of co-operation are already being defined. And there is little doubt AI systems will play a far greater role in people's lives. In the future, machines could predict errors or disruptions in work processes (for example, in the context of predictive maintenance) or conduct the beginning of a phone call with a customer.

These changes require adaptations. But who will have to adapt? Who will determine which adaptations are made and what form should they take?

Expectations placed on workers

We often hear that workers should receive continued training, to remain ‘employable’. And who has anything against training? Yet the debate isn’t progressing beyond general demands. How individual professions and sectors are affected, as well as the corresponding satisfaction of the demand for training, have not been adequately addressed or implemented. Instead, expectations are placed on workers to behave in an economically rational way—and to kindly get some training.

Participation in continued professional training programmes has in fact risen since 2010 and while now stagnating is at around 50 per cent. On-the-job training enjoys particularly high acceptance. But it is also clear that not all workers are being reached. Individuals with less formal education, those in smaller firms, those who are older or who work part-time participate less in continued training.

One study actually showed that those workers who can be easily replaced are the least likely to participate in training. In a way, we are perpetuating the inequalities of our school system. We are in danger of ending up with an even more divided labour market, with well-paid specialists on the one side and a new precariat which performs ancillary tasks—before and after the algorithm—on the other.

Contextual conditions

To ensure that current changes lead to more rather than less social cohesion, we have to create contextual conditions that make workers feel protected from unfulfillable demands. After a 40-hour working week, most of us do not have the time alongside family and care work to take part in a training programme.

This will become even less likely if labour gives into demands from the employer lobby for more flexible working times and a softening of free time. We must create more time and space to empower and protect workers who—for whatever reason—don't want this. Here legislation, such as in Germany the Qualification Opportunity Act and the Work-for-Tomorrow Act, as well as in-company guidance on continued training, should play an important role.

With these legislative initiatives, Germany's labour minister, Hubertus Heil, has already proposed or begun to enact improvements. Under discussion are the many aspects of how to finance continued training programmes. In the Work-for-Tomorrow Act, continued training is made extremely attractive for companies affected by structural change.

Workers don't just need subsidies—they need time and guidance. And the Federal Labour Office has been offering continued training advice since the beginning of 2019. A right to continued training should guarantee that employers provide enough time for it. Under such conditions, programmes can be created that empower workers.

Works councils

Works councils are key actors making sure that such opportunities are actually used. They are not 'inhibitors', trying to prevent the introduction of AI systems. Rather, they ensure that processes are implemented well and that tasks are redistributed. They do the preliminary work that will lead to greater acceptance of AI systems within companies, while at the same time providing better working conditions for the workforce.

For this to happen, we need works councils that are knowledgeable about the material. At the same time, works-council members, espe-

cially those who are not exempted from their regular tasks to take care of their works-council duties, have enough on their plate. We cannot simply unload additional tasks on to them. So they must be able to bring external expertise, on AI, data privacy and additional aspects of digitalisation, into the workplace.

Some employers are already engaged in active union-busting. In future, it will be even easier to hinder the activities of works councils if workers see and talk to each other less and less, because they will ever more frequently be working at different times and in different locations. A sense of belonging and exchange with works councils can be weakened this way.

As a reaction, work councils must also become more digital. As companies are redefined, we also need new forms of organising works councils—for example, by organising the next election via a Messenger group. We want to address such issues and others in an amendment to the Works Council Constitution Act.

Shaping the transformation

But which abilities and skills, which kinds of knowledge will broad segments of the population—such as, for example, a 56-year-old steel worker, a 32-year-old father working part-time or a 61-year-old tax adviser on a temporary contract—need in the future, to help shape the transformation and get closer to the goal of good work?

Already today, human-machine co-operation places new demands on workers. How will we deal with our algorithmic or robotic colleagues?

Complete trust in an AI system or dealing blindly with decision processes—similar to the annoying clicking away of cookie settings on websites and the acceptance of unread terms and conditions—should not be what we see in workplaces. For this reason, the

ability to think critically and to question results is all the more important.

Such ability must be founded on basic knowledge of how AI systems work. This does not mean everyone needs to be able to write code. But we should learn that AI systems and their decision-making processes have strengths and weaknesses. A successful instance of the imparting of knowledge is the ‘Elements of AI’ online course, now available in a number of EU languages, including German.

Technical knowledge also remains relevant because it empowers workers to provide critical feedback on the implementation and continued development of AI systems—because not each and every technical possibility can be practically and reliably implemented. Technical knowledge keeps workers on an equal footing with the algorithms.

Besides good formal training, tomorrow’s workplace will require from workers more communication and teamwork, because the tasks under discussion will only be able to be mastered by a group. Here, workers’ ability to give and receive constructive criticism will be key to maintaining their ability to keep learning.

What are called ‘soft skills’ today could soon prove to be ‘hard skills’, essential for companies’ success. Such skills are formed and fine-tuned through their daily application. Both employers and employees face the challenge of organising work in the future in a way that will foster these indispensable skills.

Good work

Through works councils and staff committees, workers must be integrated into the planning of continued training. This includes the development of training programmes and the form they will take, and the preparation of plans for a qualified workforce.

It is the duty of social democracy to make sure that human beings don't get the short end of the stick in the human-machine partnership. At the end of the day, machines should enable more autonomy and bring us closer to the goal of good work. We do not need another relationship of dependency in which the machine is constantly telling us where to go and which movements we should make.

We'll do that ourselves. To be able to codetermine the shape of the workplace of tomorrow, we need new contextual conditions, of which education and continued training are part—but only part.

Daniela Kolbe is a social-democratic member of the German Bundestag from Leipzig. She is a member of the labor and social affairs committee and chair of the Bundestag's study commission (*Enquete-Kommission*) on artificial intelligence.

FOUR

USING AI IN THE OFFICE FOR GOOD WORK

MARKUS HOPPE AND NADINE MÜLLER

The discussions about digitisation and artificial intelligence (AI) mostly take place from the perspective of industrial production, as is evident from the ‘Industry 4.0’ debate which dominates in Germany. By contrast, little attention has been paid to tasks involving the handling of individual cases and how they shape large parts of the service sector as well as, indirectly, industrial companies (‘white-collar work’). The ‘smartAIwork’ research project has however investigated the effects of AI in case handling and developed design solutions.

Case handling mostly involves administrative or office work. The spectrum ranges from simple data entry to complex tasks that require a high degree of creativity and knowledge, such as in information-technology development or application of legal regulations. Simple office tasks with high portions of routine work—maintaining address files, for instance—are suitable for (partial) automation by means of software and algorithms. AI, on the other hand, is used to assist people in performing demanding case-handling tasks. The aim should be to ‘relieve’ the work of monotonous, burdensome aspects and to create more space for the ‘actual’ work.

Typical applications for AI in the office include:

- ‘intelligent’ chat bots which are capable of learning in customer service, including in banks or local public transport;
- AI-supported assistants within human-resource management, or ‘AI recruiters’, and
- ‘intelligent’ robotic process automation for document management, such as for settling accounts for business trips or in procurement.

AI is currently not very widespread, however, and no more than a quarter of companies use corresponding technologies in their office work or plan to do so. Since, compared with industrial production, case handling is less easy to translate into standard processes, the opportunities for using AI in office work are limited.

‘Human factor’

This is especially so where the ‘human factor’ plays a major role—in the individuality of customer requests in banking or, more generally, where greater trust in decisions or an ability to contextualise is required. Furthermore, as the ‘smartAIwork’ project also shows, there are hurdles when it comes to the availability and quality of data for AI applications. This is a major challenge, especially for small and medium-sized enterprises.

Whether case-handling activities are replaced by AI should not however just depend on whether suitable uses can be found and whether substitution is technologically possible. There are sometimes good reasons for *not* automating certain activities. In addition to economic efficiency, these include legal restrictions, such as European Union constraints on legitimate data use.

Moreover, the combining of automated and non-automated activities in professional tasks can mean that the complexity of the tasks increases, which can increase workload. In addition, AI is only designed for a narrowly limited area of application and only shows its capabilities to their best advantage there. The inability to respond adequately to unpredictable changes in the work process outside of its defined field of application therefore places a technological limit on its use.

New interactions

AI is however expected to lead to new forms of interaction between humans and technology, which can simultaneously improve human work and increase the efficiency of work processes. The issue of AI use is thus not just one of rationalisation and automation but particularly of assisting human work, which can also lead to improved working conditions. For AI to be effective in this sense in the office, operational concepts must be designed on the basis of suitable general conditions.

The results of ‘smartAIwork’ show that the potential risks of using AI—particularly job losses and deskilling—can be avoided if certain factors are taken into account: legal and ethical standards, ergonomic findings about good work design and participative approaches to planning and implementing AI projects. The latter also help increase the extent to which AI is accepted by those employed in case handling. There is a greater chance of improving working conditions and results if AI is used as an assistant, not as a rival, to human work.

To establish the necessary general conditions and participatory processes, the support of politicians and social partners is required. They are asked to play their part to ensure that AI support in case handling leads to ‘good work’.

Ethical guidelines

In March, to mark the opening of the ‘AI Observatory’ of the Federal Ministry of Labour and Social Affairs, the German services union ver.di published ‘Ethical guidelines for the development and use of artificial intelligence (AI)’. These should serve as the basis for discussions with developers, programmers and decision-makers. Their target group also includes employees who are involved in the conception, planning, development, purchasing and use of AI systems in companies, and who therefore bear responsibility for them.

The union took a position on AI for the first time at the end of 2018, emphasising that the goals behind its development and deployment were central. AI should serve people—so the goals of, and premises for using, AI must be defined as precisely as possible. It is of the utmost importance that ‘good work by design’ is the approach from the start. To implement this, employee representation needs to be strengthened: participation needs to be ensured as early as possible during planning.

With a view to the impact AI will have on employment, we urgently need a targeted and strong commitment from politicians to establish employment relationships that have social-security protection, to strengthen the collective-bargaining system, to distribute employment fairly and to upgrade the social services required in society. A political debate is necessary concerning the areas in which AI assistance makes sense and is socially desirable. Assistance systems should also be preferred to autonomous systems, in terms of risk and workload management.

Additional training

Options for lifelong, in-service training must be established to be able to counter the rapid shift in the AI-shaped world of work from the point of view of the labour force—for example, through state-sponsored part-time work combined with continuing professional development, and a right to such additional training enshrined in a nationwide law. Ethical, social and democratic aspects need to be integrated into this education and further training, which is mostly otherwise of a technical nature only.

More binding worker protection and the safeguarding of personal rights are also required. Employee data protection is overdue, because the special dependency of employees is particularly evident in the AI context. For example, a ban on the collection and processing of biometric data from employees is urgently needed, as ‘pilot projects’ that use AI in call centres make clear. The ‘Ethical guidelines’ follow up on these positions and deepen them—particularly with a view to providing guidance and support for those who develop, introduce and use AI applications.

Markus Hoppe is a sociologist on the research staff of INPUT Consulting gGmbH in Stuttgart, focusing on the transformation of work through digitisation and AI, industrial relations and industrial sociology.

Dr Nadine Müller is head of the department ‘Innovation and Good Work’ in the ver.di federal administration in Berlin.

FIVE

MADE IN AFRICA: AFRICAN DIGITAL LABOUR IN THE VALUE CHAINS OF AI

MARK GRAHAM AND MOHAMMAD AMIR ANWAR

In discussions about the locations comprising the key productive nodes of artificial intelligence and other next-generation digital technologies, African workers rarely get a mention. Autonomous vehicles, machine-learning systems, next-generation search engines and recommendations systems—how many of these technologies are ‘made in Africa’? The answer, actually, is ‘all of them’.

In a paper from which this article is derived, we make visible the invisible and bring to light the role African workers are playing in developing such key emergent, and everyday, technologies—which underpin, or soon will, the enormous profits made by large technology companies based in the global north. In the context of hyperbolic claims about automation and robotisation—and the impending technological unemployment they are predicted to herald—human labour, including that of African workers, remains very much a part of contemporary digital capitalism.

Production networks

We conducted a five-year study (2014-19) in South Africa, Ghana, Nigeria, Uganda and Kenya, involving in-depth interviews and group discussions with more than 200 stakeholders—including workers, managers of outsourcing firms, government officials, trade unions, employment agencies, private-sector associations and industry experts. This enabled us to construct a snapshot of the key ways in which African digital labour has been integrated into the production networks of digital products and services being deployed around the world.

We focused on machine learning and digital decision-making. These activities are performed by workers employed within firms or operating as freelancers through digital work platforms (such as Upwork, Freelancer.com and Amazon Mechanical Turk), which act as intermediaries between employers and workers in a planetary labour market. Much hides behind the sleek, automated surfaces.

African workers play an important role in building and maintaining these technologies—acting as ‘data janitors’. Real people are still needed to structure, classify and tag an enormous amount of unstructured information for companies using machine-learning algorithms in their products.

While many scholars are predicting that machines will replace humans in the production process, thus increasing unemployment around the world, automation is not always what it seems. Technological advances and use of machines in production can destroy jobs in one location (primarily richer regions), yet can also open up many lower-income work possibilities for workers in poorer countries.

Once we acknowledge that many contemporary digital technologies rely on a lot of human labour to drive their interfaces, we can begin to piece together what the new global division of labour for digital

work looks like. We need detailed empirical studies of where value is created and captured in these production networks, which are opaque by design. Research can start to make the invisible nodes of these chains more visible and highlight the pay and working conditions of the workers who make everything possible.

This is not to say that many of these digital workers are poorly paid by local standards, or that they are ungrateful for their jobs. But high unemployment and a large informal sector mean these digital jobs receive overly positive reviews, while the risks are sidelined. And digital workers in Africa are still earning only a tiny fraction of the profits generated from their labour.

Socio-political response

The contemporary digital economy thus offers jobs and opportunities to African workers, but even more of an opportunity to the international corporations which seek to profit from their labour. There is no easy means for firms and individuals based in the world's economic margins to move up global value chains, but this does not mean that we should throw up our hands and accept the *status quo*.

As digital connectivity spreads to the last corners of the world, we hope this knowledge will help build a greater socio-political response to the relatively labour-intensive nature of the contemporary digital economy, in which African workers play a significant role in value creation. Once we acknowledge that many contemporary digital technologies rely on a lot of human labour to drive their interfaces, we can begin to piece together what the new global division of labour for digital work looks like and aim—at both the global and local scales—to make some of these value chains more transparent, ethical, and rewarding.

Meantime, there is still much to be done to understand better African digital labour, to challenge labour processes and employment relations, to improve the quality of work and to identify the common interests of workers—and the ways their labour connects distant sites of production and consumption.

Mark Graham is professor of internet geography at the Oxford Internet Institute and director of the Fairwork Foundation.

Mohammad Amir Anwar is a lecturer in African studies and international development at the Centre of African Studies, University of Edinburgh.

SIX

CAPITALISM'S MIRROR STAGE: ARTIFICIAL
INTELLIGENCE AND THE QUANTIFIED
WORKER

PHOEBE MOORE

Control panels are the obvious place to run operations centrally. The control rooms of Star Trek's fantastical *Enterprise* (and the hub of the actual Project Cybersyn under Chile's radical president Salvador Allende) in the 1960s and 70s were however operated by humans with relatively primitive technologies.

Today, much of the work of the people we imagined in these rooms—the bouffanted women in silver A-line dresses and men in blue boiler suits pushing buttons to operate the manoeuvres of galactical imperialism—is done by computers. But what will happen when the proverbial windows looking out to the galaxies only display a cadre of robots and the control panels' blinking lights are the only reflective glimmer?

So-called Industries 2.0-4.0 have seen an onslaught of machines and machinic competences in the workplace control rooms of today, via robotic process automation, semi-automation, machine learning and algorithmic management systems. Digitalised workplace design and surveillance techniques are oriented around the rise in new technologies, where the processing and quantification of workers' data is seen to be necessary for a company's competitiveness.

People analytics

The contingent technology for workplace processes to reach a new pinnacle of computational sophistication is the rise in artificial-intelligence tools and applications. AI allows semi-automation of decision-making processes via machine learning, which is particularly applicable in the case of human-resource driven ‘people analytics’ (PA), where predictions and prescriptions about job candidates and workers—or ‘data subjects’ as the General Data Protection Regulation (GDPR) puts it—can now be made based on quantification techniques applied to data sets.

Put simply, with the use of PA, we are asking machines to relay truths, or subjective images about other people, via computation. While we once expected the machine to mirror the human, we now seem to be looking into a machinic mirror for our own reflection and those of others. The full implications of this ‘mirror stage’ of capitalism—to borrow a phrase from the psychoanalyst Jacques Lacan—are yet to be played out but are exceedingly important.

For Lacan, the mirror stage was the moment in which the child realises her separation from the rest of her environment. The mirror stage for what I am calling ‘smart workers’ within capitalism today must be a moment of defying the assumption that we are inexorably subsumed into a machinic subject, retaining the firm scaffolding of what makes us human and posing resistance to a purportedly automatic domination. Given growing expectations that AI will become universal, to avoid the most negative implications it implies for workplaces and workers with regards to automation and surveillance, it is increasingly important to exercise reflexivity and retain our human autonomy, as decision-making about workers is increasingly based on quantification and automation.

Machine learning

People analytics is perhaps the best-known form of AI-augmented workplace tool. Generally speaking, PA is a set of human-resource (HR) activities which rely on a process whereby managers can identify patterns and compare them across data sets collected about workers.

The AI component in PA lies in how algorithms are set up to make the decisions, via machine-learning procedures. Big data, algorithms and machine learning are central in digitalised recruitment, where decisions about talent spotting, interviewing, leadership prediction, individual worker performance, health patterns across workers and other operational management issues can be digitally assisted.

Indeed, machines become the mirror for workers' subjectivities via quantification. Predictions are made about applicants regarding aptitude and job fit—and, once workers are in position, many things can be assessed, ranging from the diligence of their work to their likelihood for disengagement.

A Deloitte report indicates that 71 per cent of international companies have reported they value PA and see it as a priority, because it allows management to conduct 'real-time analytics at the point of need in the business process ... [and] allows for a deeper understanding of issues and actionable insights for the business' to deal with what have been called 'people issues'. In other HR-related reports, the revelations of 'people risks' and 'people problems' which PA can unveil throw the concept of the mirror phase of capitalism into sharp relief: who are we (humans), in the machine's reflection?

Increased stress

PA is likely to increase workers' stress if data are used in appraisals and performance management without due diligence in process and implementation, leading to complaints about micromanagement and feeling spied on. If workers know their data are being read for talent spotting or deciding possible layoffs, they may feel pressurised to advance their performance, and begin to overwork, posing significant risks. Another risk arises with liability, where companies' claims about predictive capacities may later be queried for accuracy or personnel departments held accountable for discrimination.

Indeed, if algorithmic decision-making in PA does not involve human intervention and ethical considerations, this HR tool could expose workers to heightened structural, physical and psychosocial risks and stress. How can workers be sure decisions are being made fairly, accurately and honestly, if they do not have access to the data held and used by their employer? This should be dealt with to some extent in the European Union context with the GDPR but that is by no means a *fait accompli*.

PA practices are particularly worrying if they lead to workplace restructuring, job replacement, job-description changes and the like. In any case, the use of machine learning to make predictions and provide analyses about people relies on specific kinds of intelligences prioritised under capitalism—efficiency, reliability, competitiveness and other data-driven imperatives—which may or may not reflect who individuals are, or would like to be, in modern society.

Research necessary

Many high-level governmental and organisational reports are predicting that AI will improve productivity, enhance economic growth and lead to prosperity for all—in a similar way 'scientific

management' was once heralded. As with scientific management, however, high-level discussions do not seem to link the anticipated prosperity directly with the realities of the everyday (and everynight) human work which ultimately fuels growth. Meanwhile, various AI-augmented tools and applications are being introduced to improve productivity, in factories and offices and 'gig' work.

There is a lot of research on *automation* but not on how AI, as a form of semi-automation, carves out the capacity for substitution of human activities in the workplace. There is also extensive research on *surveillance*, but again not scrutinising how AI facilitates advances in surveillance in the workplace.

Scholarly and governmental research on these subjects should take AI seriously by putting a metaphorical mirror into place for social reflection about how these processes occur and on which assumptions they rest—rather than presenting AI merely as forms of autonomous software and immutable techniques for facilitation. While there have been significant inroads in climate, medical, fashion, insurance and justice-systems research, studies on AI's uses to evaluate workers and aptitudes through quantification are lagging behind. Stories of discrimination and bias are already making headline news where PA has been applied and, without reflection on the mistakes made in AI and quantified analyses of workers, this is set to continue and even get worse.

Digital democracy

The rise in data accumulation in recent times and the reliance on algorithms for workplace decisions has led to the possible removal of the role of the physical manager through a machinic system. If workers were to take over workplace control rooms through deciding which tools and processes are applied, digital democracy at work could be imagined.

But the use of AI undemocratically could just as easily occur and lead to the removal altogether of human autonomy, via automation, from workplace decision-making and tasks. The current Covid-19 crisis has also led to the rise in online working, giving increased leeway for quantified judgements and machinic management.

More research is needed in these areas, to get a full picture of what AI will mean and, in many cases, already means for human-machine relations in workplaces. What precisely are the types of intelligences which we expect today from machines and are these really reflective of human intelligence? Why do we choose the categories of intelligence that we do, and how are data collection and processing activities relevant to the affective side of the human experience?

Perhaps most importantly, what are the surrounding risks for workers as technology advances and as we begin to question our own role in production and think about that of the machine, as AI is set to increase *its* autonomy? The question more broadly for humanity is: who do we think we are as we reach the mirror stage in capitalism, where we should realise we are separate and retain autonomy from a machinic subject?

As we busily instal machines into workplaces via robotics and management tools with seemingly superior intelligence to ourselves, we should ask: in whose (or which) reflection are we now looking?

Phoebe Moore is associate professor in political economy and technology in the School of Business at the University of Leicester and director of its Centre for Philosophy and Political Economy.

SEVEN
DESIGNING AI TOOLS TO BENEFIT
WORKERS

FLORIAN BUTOLLO

The discourse on artificial intelligence and work is shaped by conflicting narratives. Disempowering notions about mass unemployment and a loss of human control in the face of ever-more-powerful machines are widespread. But AI also inspires visions of human empowerment, according to which labour will be upgraded as machines support human effort and relieve us from the burden of onerous work, leaving us with more interesting, creative and cognitive tasks.

Both narratives are one-sided, deriving projections as to the future of work from the nature of technology as such. To overcome this simplistic dichotomy, the social context in which AI is introduced needs to be addressed. It is not just an interaction between man (or woman) and machine—AI is implemented within a far-flung division of labour, which entails multiple forms of co-operation, task specialisation and inequality. To answer the question of who benefits and who loses through its introduction, it is thus necessary to ask how relations of power between human agents are reconfigured.

Significant limitations

Hubris surrounds the term AI and is responsible for many of the misconceptions. The present technological path of machine learning has generated astonishing breakthroughs, yet significant limitations are encountered when the calculated results are contextualised and applied.

And while it is now possible to detect patterns in massive data sets which surpass the capabilities of human reason—essentially amounting to a different form of intelligence than that of humans—the ‘predictions’ derived from these are structurally conservative. They merely project such patterns into the future, based on correlations established rather than a deeper understanding of underlying factors.

What is more, AI systems continue to be trained towards very specific tasks and cannot transfer capacities to different data sets or changed surroundings. In other words, AI delivers highly-sophisticated statistical evidence for processes of high regularity in controlled surroundings.

There is a multiplicity of applications where these forms of pattern recognition matter, especially in the image or speech recognition and match-making which constitute the main fields of AI today. But this is intelligence in the statistical sense, not anything equivalent to human intelligence.

It fails to work once there are more complex, multi-factor environments involved—think Brexit or the notorious butterfly which might trigger a hurricane in a different region of the world! Human reasoning must step in to contextualise AI results, to understand its implications in real-life scenarios.

Augmented intelligence

In terms of possible impacts on work, this means AI can be used to subordinate workers to the mechanical calculations of the machine or to empower them to contextualise and use AI as *augmented* human intelligence. Both approaches exist.

The first path isolates the work process from its real-life context. The design of a logistics warehouse or simple manufacturing operation can easily be translated into a data model with input, processing and output variables. AI algorithms can recurrently recalculate the set of factors involved and transmit these to human agents, obliged to follow suit.

Such forms of automated decision-making leave little room for the opinions of workers. Devices displaying the next operation approximate to ‘objective’ efficiency and functionality, to the extent that it becomes futile to argue. The bugs and readjustments that (as always) occur remain the preoccupation of data scientists and management. Workers are supported in their actions but they become highly replaceable, their bargaining power undermined.

The second path ascribes the tasks of contextualising AI to workers. AI might provide transparency about the current state of processes and hints as to possible measures to smooth the operation of a firm, be it a factory or an office. Yet humans face the challenge to interpret such results, based on their capacity to assess the surrounding factors and their experience. This way, decisions can be augmented via a translation and adaption to real-life conditions, building on work experience, intuition and general reasoning. These capacities can be developed through enhancing workers’ capacities to understand, interpret and act upon automated decision-making.

New forms of interaction

It is easy from this to deduce scenarios of a downgrading or an upgrading of work. The point, however, is to identify the variables that affect whether one tendency or the other predominates. This is not rooted in the structural surroundings of certain work contexts or in technology itself but in the active design of new forms of man-machine interaction.

Three dimensions are particularly relevant. The first concerns the fundamental question of investment in technologies, the second the design of interfaces between AI and its users and the third the challenge of equipping workers to upgrade their skills.

Regarding investments, AI can be used for a broad variety of tasks which can be detrimental or supportive when it comes to workers' empowerment. The question of how technological choices affect power relations in the workplace is a complicated one which needs to move centre-stage in discussions among workers' representatives. It is linked to management choices favouring the design of enterprises as learning organisms (thus requiring the input of workers) as against neo-Taylorist options that reduce workers to narrowly-circumscribed functions.

Next, the design of technology becomes an important matter for workplace politics. Do the interfaces of AI systems indicate a set of options and the contingency of automatically-generated results? Or do they narrowly prescribe actions that will be mistakenly taken as givens by human agents? Does AI challenge us to interpret its results or relegate us to an observing position? These are delicate questions as to what roles are ascribed to workers in AI models.

Finally, how do companies support workers in developing new skills in a setting of augmented intelligence and how is this incentivised?

Calls for more extended training and lifelong learning are widespread—workers need to acquire a deeper understanding of automated processes to make the right decisions, involving the skills to negotiate the translation of insights from the data level to physical processes and real-life communication.

But if workers need to learn more and constantly, how is this to be encouraged? If lifelong learning becomes a requirement that is not compensated through higher wages and relief from other responsibilities, it could soon become not a blessing but a burden. Workers would need to run to stand still in the hierarchies of the workplace.

Tough challenges

All these dimensions constitute tough challenges for workers, works councils and trade unions. They are relevant fields for designing the workplaces of the future, as the technological choices and their embeddedness are surrounded by conflicting interests, in which workers need to strengthen their voice. This necessitates an upgrading of the side of labour towards stronger capabilities in evaluating technologies and putting them to use according to their interests.

And this challenge becomes enduring: AI systems are not merely another machine which once introduced keeps on working in the same way, but learning organisms which modify their functions going forward. AI thus requires an augmentation of bargaining intelligence, so as to be capable of affecting the balance of forces on the shopfloor to workers' advantage.

Dr. Florian Butollo is a researcher at Berlin Social Science Center and head of the research group 'working in highly-automated digi-

tal-hybrid processes' at the Weizenbaum Institute for the Networked Society in Berlin. He is an adviser to the study group on AI in the *Bundestag*.

EIGHT
CONTROLLING THE EFFECTS OF AI ON
WORK AND INEQUALITY

CHRISTIAN KELLERMANN AND MAREIKE WINKLER

What will be the effects of the digital transformation on jobs? Job creation outnumbering digital job destruction is part and parcel of standard artificial-intelligence (AI) prophecy. But the extent to which work tasks are upgraded—rather than downgraded or even replaced—is determined by at least two dimensions: the technical side and the work aspect.

Today, in the production and service sectors ‘digitalisation’ in most cases means the use of smartphones and tablets. These devices undoubtedly are operated by complex technology—such as AI. But full automation is not yet the main reality.

Nevertheless, the robot—another smart device—is already replacing human work, which has negative effects on wages. Middle- and low-skilled jobs in particular have been affected by information and communication technology (ICT) and robots since the 1970s and 80s. The consequences are decreasing wages on the one hand, and productivity growth and rising ‘digital dividends’ on the other. These dividends, however, are mainly received by the capital owner and explain (in part) the shrinking wage share.

In a country such as Germany, robots are certainly common but in industry they are very concentrated, especially in automotive manufacturing. The vast majority of studies therefore conclude that digitalisation drives the automation of work tasks in certain domains, but also creates much—or even more—work in other, less automated areas, primarily in the service sector.

Daring assumption

Believing that digitalisation must have automatic positive effects on total employment, however, would be quite daring. It depends on the assumption that demand for work lost is (over)compensated by new demand for work elsewhere.

The more precisely this presumed multiplier effect is broken down, the more pronounced the doubts about the associated technology optimism become. The promise is that sectoral productivity gains through digitalisation lead to ‘prosperity for all 4.0’. Yet not only have such ‘trickle-down’ claims gone through a credibility crisis in the last 30 years; they also present a very demanding scenario when it comes to digitalisation.

On the one hand, the assumption is correct that demand for services—or, put more generally, for manual tasks—will increase if some employees receive higher wages because they benefit from digitalisation. On the other hand, these tasks are relatively price-inelastic, so if their price falls due to the use of technology, demand for them will not grow to the same extent.

Technology will not automatically lead to a general increase in prosperity. Instead of focusing on the side of technology and associated investments, a social technology assessment is required, in which the distributional effects of digitalisation are carefully considered.

Without controlling AI's differential effects on the labour market, inequality will continue to rise.

Tacit knowledge

Luckily, the scenario of a highly automated industry remains a vision for the future, mainly because of the complexity of even simple work. Each job comprises a whole bundle of experiences—no matter how routinised the tasks may be. The capacity to work generally requires tacit knowledge about how to deal not only with complexity but also uncertainty, which is out of reach for 'tool' or special-purpose AI.

Today, so-called 'world knowledge' can be formalised in simple individual cases in AI models, but it is expensive, resource-consuming and always reductionist. The marginal utility of today's AI is still very limited and does not justify scenarios of massive job losses. These assumptions are usually based on a simplistic understanding of routine work and the production process.

When it comes to regulation, one of the most urgent issues is thus to counter the digital anxiety of many workers with a realistic assessment and an appreciation of their individual working abilities. Practical, including technical, co-determination is also needed in the digitalisation of operational processes.

This requires the strengthening and extension of co-determination structures and rights. Co-determination serves here not only to control the technology but can also be a supportive factor in investment decision-making, which often is not properly recognised by management alone.

Prosperity for all

Finally, a forward-looking policy has the responsibility to correct potential, excessive and unequal distribution effects—so that eventually prosperity for all is in fact created. In the short term, redistributive measures are essential to pursue a social ‘Pareto optimum 4.0’; in the long run, a transition plan is needed towards a world of work which tames advanced AI.

Such shared prosperity will be largely material in nature. But it can also be increasingly immaterial—including a reduction of working time.

Dr Christian Kellermann is managing director of *Das Institut für die Geschichte und Zukunft der Arbeit* (IGZA), the Institute for the History and Future of Work.

Mareike Winkler is research assistant at the IGZA. The opinions expressed are those of the authors alone.

NINE
AI: THOSE ARE CITIZENS MARCHING, NOT
ROBOTS

MIAPETRA KUMPULA-NATRI

We are witnessing another industrial revolution—a digital one. Rapidly evolving technology, superfast connections such as 5G, the massive amount of data this connectivity generates and artificial intelligence will reshape the lives and societies we know today.

Globally, the total amount of data is doubling every 18 months. In other words, in 2019 we were using only 1 per cent of the data which will be in use by 2030. This creates yet unimaginable possibilities for innovations, new business models and services.

Yet who will this trend benefit? Will the pool of data be used to build a human-centric digital society or could it end up concentrated in the hands of a few global actors, benefiting only the already wealthy?

The digital revolution should neither leave anybody behind nor lead to a ‘race to the bottom’ with regard to labour and social standards. Everybody must be included. We must not stifle innovation but data usage cannot be an unregulated vacuum. We must empower citizens to have better control over their data and use data as a tool to benefit people and societies as a whole. As legislators, it is our task to

establish a regulatory framework that promote an inclusive, human-centric data economy in Europe.

AI has been a clear priority for the current European Commission from day one. But it was the commissioner for the internal market, Thierry Breton, who really put the emphasis on data. Data and AI go together: if we do not have data ‘flowing’ between different actors, whether public or private, and across borders, Europe cannot be number one in the world in reaping the benefits of digitalisation or AI.

On the European Parliament’s own-initiative report on data strategy—its answer to the commission communication in February—I have the honour to act as the industry committee’s *rapporteur*. The aim is to find a parliament position before the commission publishes concrete legislative proposals, such as the envisaged enabling legislative framework for the governance of common European data spaces, data act and implementing act on high-value data sets. From the standpoint of European citizens, the focus is clear: how to harness the potential of data to enable new services, business opportunities and jobs, while ensuring the digital transformation doesn’t leave behind common European values?

At the same time, it is important to understand that the digital market is truly a global one. I have an opportunity to follow also the global digital debate from the international-trade perspective as a standing rapporteur on World Trade Organization e-commerce negotiations in the European Parliament’s international trade committee. The EU must be an active global player and influence the development of the digital world based on its values—not the other way around. For example, we must put the focus on European competition policy: Europeans must define the rules, values and level playing field of the market; we should not be satisfied only with what others dictate.

Trust needed

Building a human-centric data economy and human-centric artificial intelligence starts from the user. First, we need trust. We need to demystify the data economy and AI: people tend to avoid, resist or even fear developments they do not fully understand.

Education plays a crucial role in shaping this understanding and in making digitalisation inclusive. Although better services—such as services used remotely—make life easier also outside cities, the benefits of digitalisation have so far mostly accrued to an educated fragment of citizens in urban metropolises and one of the biggest obstacles to the digital shift is lack of awareness of new possibilities and skills.

We need action throughout Europe, all the way down to the local level, to give our citizens the tools to understand rapid technological change—as well as investing in new engineers, software developers and visionaries via our education systems, reskilling and lifelong learning. How can employees and small and medium enterprises be innovative, if they do not have the knowledge?

Exemplary initiative

An exemplary initiative is a Finnish-developed, free online course, ‘Elements of AI’. This started as a course for students in the University of Helsinki but its wider potential was soon realised and the paradigm changed: the new aim of the university and its partner company was to educate 1 per cent of Finnish citizens in the basics of AI. The course boomed and the goal was reached in no time among Finland’s 5.5 million population.

Finland held the presidency of the Council of the EU during the second half of 2019. In a departure from tradition, it did not give

any gifts during the presidency, expect one—extending the goal to offer basic knowledge of AI to 1 per cent of all European citizens. In co-operation with the commission, the course will soon be available in all official European Union languages.

So far, more than 430,000 people from over 160 countries have taken the course. It is not designed only for professionals or digital ‘nerds’ but for common people: the only requirement is an internet connection and a will to learn. The course is digital education and lifelong learning *par excellence*. It’s a concrete and easy-to-use initiative which really has a multi-functional purpose—you can use it just to learn the basics of AI on your own from your bed in the evening, or take the course as a part of the education system in school, university or work. It is already part of the curriculum in almost every Finnish university and some employers in Finland have advised their employees to take it—just to keep up with the evolving world.

Gender balance

Another key issue is gender balance. AI learns from real-life data and there is a tangible risk that it will adopt existing biases and even make them more apparent. This is why the coders and users of AI-based technology need to be diverse. Yet how long have we talked about the small number of women in the technology industry? I graduated as an engineer in the 1990s and that topic is certainly not new.

Concrete possibilities for equal participation make the world more balanced. In the Nordic countries, the majority of participants on the ‘Elements of AI’ course are female and in the rest of the world the proportion exceeds 40 per cent—more than three times as high as the average ratio of women working in the technology sector. After the course had been running in Finland for a while, the

number of women applying to study computer science in the University of Helsinki increased by 80 per cent.

Let's be inspired by this and relentlessly continue our work, from the grass roots to the global level, to ensure we build fair, equal and progressive digital societies.

Miapetra Kumpula-Natri is a member of the European Parliament, representing Finland and the Socialists and Democrats group, and of its Committee on Industry, Research and Energy. For 11 years, she was a member of the Finnish parliament.

TEN

EXPLAINING ARTIFICIAL INTELLIGENCE IN
HUMAN-CENTRED TERMS

MARTIN SCHÜSSLER

Intelligent systems, based on machine learning, are penetrating many aspects of our society. They span a large variety of applications—from the seemingly harmless automation of micro-tasks, such as the suggestion of synonymous phrases in text editors, to more contestable uses, such as in jail-or-release decisions, anticipating child-services interventions, predictive policing and many others.

Researchers have shown that for some tasks, such as lung-cancer screening, intelligent systems are capable of outperforming humans. In many other cases, however, they have not lived up to exaggerated expectations. Indeed in some, severe harm has eventuated—well-known examples are the COMPAS system used in some US states to predict reoffending, held to be racially-biased (although that study was methodologically criticised), and several fatalities involving Tesla’s autopilot.

Black boxes

Ensuring that intelligent systems adhere to human values is often hindered by the fact that many are perceived as black boxes—they thus elude human understanding, which can be a significant barrier for their adoption and safe deployment. Over recent years there has been increasing public pressure for intelligent systems ‘to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made’. It has even been debated whether explanations of automated systems might be legally required.

Explainable artificial intelligence (XAI) is an umbrella term which covers research methods and techniques that try to achieve this goal. An explanation can be seen as a process, as well as a product: it describes the cognitive process of identifying causes of an event. At the same time, it is often a social process between an explainer (sender of an explanation) and an explainee (receiver of an explanation), with the goal to transfer knowledge.

Much work on XAI is centred on what is *technically possible* to explain and explanations usually cater for AI experts. But this has been aptly characterised as ‘the inmates running the asylum’, because many stakeholders are left out of the loop. While it is important that researchers and data scientists are able to investigate their models, so that they can verify that they generalise and behave as intended—a goal far from being achieved—many other situations may require explanations of intelligent systems, and to many others.

Many intelligent systems will not replace human occupations entirely—the fear of full automation and eradication of jobs is as old as the idea of AI itself. Instead, they will automate specific tasks previously undertaken (semi-)manually. Consequently, the interaction of humans with intelligent systems will be much more

commonplace. Human input and human understanding are prerequisites for the creation of intelligent systems and the unfolding of their full potential.

Human-centred questions

So we must take a step back and ask more values- and human-centred questions. What explanations do we need as a society? Who needs those explanations? In what context is interpretability a requirement? What are the legal grounds to demand an explanation?

We also need to consider the actors and stakeholders in XAI. A loan applicant requires a different explanation than a doctor in an intensive-care unit. A politician introducing a decision-support system for a public-policy problem should receive different explanations than a police officer planning a patrol with a predictive-policing tool. Yet what incentive does a model provider have to provide a convincing, trust-enhancing justification, rather than a merely accurate account?

As these open questions show, there are countless opportunities for non-technical disciplines to contribute to XAI. There is however little such collaboration, though much potential. For example, participatory design is well equipped to create intelligent systems in a way that takes the needs of various stakeholders into account, without requiring them to be data-literate. And the methods of social science are well suited to develop a deeper understanding of the context, actors and stakeholders involved in providing and perceiving explanations.

Evaluating explanations

A specific instance where disciplines need to collaborate to arrive at practically applicable scientific findings is the evaluation of explana-

tion techniques themselves. Many have not been evaluated and most of the evaluations which have been conducted have been functional or technical, which is problematic because most scholars agree that ‘there is no formal definition of a correct or best explanation’.

At the same time, the conduct of human-grounded evaluations is challenging because no best practices yet exist. The few existing studies have often found surprising results, which emphasises their importance.

One study discovered that explanations led to a decrease in perceived system performance—perhaps because they disillusioned users who came to understand that the system was not making its predictions in an ‘intelligent’ manner, even though these were accurate. In the same vein, a study conducted by the author indicated that salience maps—a popular and heavily marketed technique for explaining image classification—provided very limited help for participants to anticipate classification decisions by the system.

Many more studies will be necessary to assess the practical effectiveness of explanation techniques. Yet it is very challenging to conduct such studies, as they need to be informed by real-world uses and the needs of actual stakeholders. These human-centered dimensions remain underexplored. The need for such scientific insight is yet another reason why we should not leave XAI research to technical scholars alone.

Martin Schübler is a Phd candidate at TU Berlin, working at the interdisciplinary Weizenbaum Institute.

ELEVEN
ARTIFICIAL INTELLIGENCE, HEALTHCARE
AND THE PANDEMIC

SELIN SAYEK BÖKE

Our world has been shaken by the Covid-19 pandemic, pushing policy-makers to scramble for solutions. And even though the full set of such solutions remains elusive, already a return to normal is debated.

But what will this ‘normal’ be? Powerful forces presume that the world before Covid-19 is the normal to which to return and it falls on progressives to push for new fundamentals—to help formulate a ‘new’ normal. Clearly this is multifaceted and one facet is the role of technology.

Undeniable role

Artificial intelligence, as a revolutionary force in restructuring production and consumption patterns, has long been on the agenda of policy-makers. The role of AI, as a creative but disruptive process in the job market, in healthcare, in education—even in shaping our democracies—is undeniable.

Given the health focus of the continuing crisis, overcoming the regulatory, ethical and medical challenges posed by the use of AI in

healthcare must be a priority. Defining the framework to do so will be a pivotal initial step in guaranteeing that the new normal produces a fair outcome—that fundamental rights are safeguarded while simultaneously improving healthcare for all.

If supported by adequate and effective regulation, AI promises a wide array of opportunities to improve public health as well as the quality and efficiency of the healthcare sector. Without such a framework, AI has the potential to be just another instrument in a system where rights are sidelined for profit maximisation and biases are reproduced systemically.

The Parliamentary Assembly of the Council of Europe (PACE) is preparing a number of reports on the implications of AI. As *rappor- teur* on AI in healthcare, I must point to existing Council of Europe legal instruments—such as the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (the Oviedo convention) and the Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data—as guides for national regulatory efforts.

Tracking and tracing

Clearly AI has played a critical role in the initial detection of the pandemic. It has been used in tracking the spread of disease and hospital capacity, in identifying high-risk patients and in developing drugs and, potentially, a vaccine. Maybe the most visible public debate regarding AI in healthcare has been over ‘testing and tracing’ apps, which have been claimed as important tools to control the spread of the virus and provide valuable information to design strategies for exit from lockdown.

AI's highly promising potential for the future of public health in Europe is however not the only reality which the pandemic has laid bare. It has offered a stark reminder of socio-economic inequalities —of the need to restrain over-marketisation and regulate markets, and to govern potential conflicts between ethical principles and market forces.

The lasting legacy of neoliberalism is manifested most notably in privatised healthcare and highly precarious job markets. This has aggravated the consequences of the pandemic, particularly for working people, for the unemployed and for the precariat. The unequal social and economic structures established and reinforced under neoliberal hegemony impede our capacity to address the challenges it has thrown up.

Equally, had there been a trusted and well-defined regulatory framework, maybe AI could have had a much larger positive impact on the coronavirus crisis. The public's concern regarding the misuse and abuse of data by states, as well as the private sector, would have been mitigated.

Totalitarian drift

We need to set a new framework capable of creating social benefits from AI while safeguarding fundamental rights and democratic governance and ensuring equality. These questions fit snugly into the debate as to what the 'new' normal will be: will the means of surveillance for the sake of health purposes accelerate a totalitarian drift or will they be governed by an empowered citizenry? And will isolationist reflexes deepen or will multilateralism, co-operation and solidarity rise to the challenge?

These questions are relevant to any discussion of AI and healthcare —the former to a regulatory framework that will ensure protection

of human rights, the latter to whether AI in healthcare will be driven by co-operation and solidarity or, in their absence, profit-seeking objectives.

Evidently, health and personal privacy can never be alternatives—they must go hand in hand. Public trust in the state and the private sector can only prevail if all their agents guarantee basic human rights in developing and using AI.

Given the urgency of doing so in the struggle against the coronavirus, it is of utmost importance to agree on at least a workable basic framework that will enhance trust and make AI operational for the better. And the Covid-19 outbreak has shed light on its critical aspects.

Empowering citizens

Such a framework should ensure that AI in healthcare empowers citizens in making better-informed decisions and provides information to hold governments accountable for the decisions they make. So that AI does not become instrumental in aggravating inequalities, it should also ensure that data and algorithms are unbiased, and that processes are transparent and inclusive.

It should be based on well-defined liability and a well-balanced public-private dialogue. It should put in place the conditions and guarantees to ensure that pursuing the collective interest does not override individual rights. It should require that technology used for monitoring and tracking is only used temporarily and does not become a permanent feature.

When the new regulatory framework is designed, the point of departure should be recognition of access to healthcare and protection of personal data and privacy as fundamental, indispensable rights. Technology-driven opportunities such as AI should be incor-

porated into healthcare systems in ways that guarantees equal access while safeguarding those rights. Only then will we not only overcome this pandemic but ensure we are ready to tackle the next one better.

Selin Sayek Böke is a Republican People's Party (CHP) member of parliament of the Turkish Grand National Assembly, representing İzmir. She is a member of the Parliamentary Assembly of Council of Europe, first vice-chair of its Socialists, Democrats and Greens group and chair of its sub-committee on the European Social Charter. With a PhD in economics from Duke University, she previously held assistant, visiting and associate professor roles respectively at the universities of Bentley, Georgetown and Bilkent and worked for the International Monetary Fund and with the World Bank.

TWELVE
A EUROPEAN WAY TOWARDS
SUSTAINABLE AI

REINHARD MESSERSCHMIDT AND STEFAN
ULLRICH

Artificial intelligence is both praised as a general solution to the most pressing social problems and loathed as a main cause of precisely these. In current debates, the focus is mostly on the ‘intelligence’ part, which is misleading because the main moral and political implications stem from the fact that AI is ‘artificial’—a socio-technical artefact.

Since the 1950s, major socio-economic and earth-system trends have followed an exponential growth pattern, called the ‘great acceleration’. Digitalisation, as a major innovation trend of the last decades, follows the same pattern, most famously observable in the computer-chip manufacturing industry as the so-called ‘Moore’s law’ (neither a law nor without limits). The exponential growth of data and computational power places higher demands on people and resources in all steps of the process we now call ‘digitalisation’ and this applies especially to AI-based systems.

Modern hardware needs a variety of raw materials, including coltan mined and processed under conditions which are socially unsustainable. Whole regions of the world (mostly in the global south) are being transformed into the ugly flip-side of the brave

new digital world. With respect to ecological sustainability, the energy needed for the extraction, processing and shipping of the components, as also for the operation of modern computer systems of Big Data and AI, is quite substantial. The rule of thumb is that a contemporary data centre needs for its operation as much electricity as a small town, mostly for cooling (life-cycle costs not included).

Nevertheless, we need information and communication technology (ICT) for the European energy revolution to happen. It can help us save energy and resources in other fields, such as mobility or electricity consumption in households—the overall energy balance depends on the aim and the motivation.

Socio-technical system

It is strange that technicians are reminded often to put the human in the loop—the human was never outside. It is humans who are creating technology, it is humans who are using technology and it is the human part of the socio-technical system AI that provides the intelligence. Consequently, ‘AI does not make us more “intelligent”, only more computationally powerful.’

And while it is tempting for a technological civilisation to seek technical solutions to all of its problems, regardless of powerful tools not every problem can be tackled by technology. Unless we change the underlying social conditions, digitalisation will increase the problems we want to solve and create additional ones.

In line with the 17 United Nations Sustainable Development Goals, the ‘ultimate goal of technology’ would be ‘to improve the human condition in a sustainable way for all of us and for our environment’. But even if this responsible understanding of innovation would become a global standard, it does not protect us from the

unintended consequences which create new problems or path-dependencies when trying to solve old ones.

Norbert Wiener, who defined cybernetics in 1948 as the scientific study of control and communication in the animal and the machine, already knew that ‘we had better be quite sure that the purpose put into the machine is the purpose which we really desire’. But that does not answer what this purpose is and who ‘we’ are—two questions better asked right at the beginning, if innovation in AI is to be safe, trustworthy, reliable and sustainable.

Therefore, political action is needed beyond the digital sphere and that leads us to the non-computable question: in what type of future society do we want to live? We need public deliberation about that, independent of putative technical ‘necessities’. In the long run, ‘any development that does not boost trustworthiness will ultimately not succeed’.

Big data-based AI calculations which are ‘good enough’ for ethically and epistemologically questionable business models cost large amounts of energy and are typically not trustworthy, for instance due to biased training datasets or machines that only pretend to learn, ‘which puts a question mark to the current broad and sometimes rather unreflected usage ... in all application domains in industry and in the sciences’. Think, for instance, of determining thus the creditworthiness of individuals.

Infrastructure lacking

Instead of primarily using AI for tracking users to personalise advertisements, the networked society of today lacks an infrastructure designed to enhance ‘individual inclusion, personal development, environmental protection, fair competition and a functioning digital public sphere’, as well as ‘access to data and services such as cloud

services, mobility platforms or a search index’—in other words, ‘the common good’. The global ‘free’ market and its powerful big-tech companies will not provide such an infrastructure, unless there is a requirement to change unsustainable business models.

It will neither emerge from the ‘move fast and break things’, surveillance-capitalist model of Silicon Valley, nor will China’s mass-surveillance state capitalism be compatible with an open, emancipatory, digital-commons ICT infrastructure. Consequently, there is an urgent need for a ‘European way’ towards sustainable digitalisation, based on trust, responsibility and public ICT.

Trust as a building block also means ensuring good engineering practices, regulation by law and a basic digital literacy. Technically, transparency and explicability play a central role. But, understood as *socio*-technical system, if AI is really to become a base technology for further sustainable innovation it must be accessible to everyone and made for the people in the common interest.

A ‘European public open space’ would provide a platform to discuss what this common interest looks like—this project for conceptualising a European public sphere is as yet only a vision but, embedded in an ecosystem of public ICT platforms, it could be a good start. Digital infrastructures which play a key role in everyday life should not be designed in favour of ‘surveillance capitalism’ and ‘networks of control’ which get more powerful with more data. Concerning web indices as fundamental infrastructure for search engines, projects such as the Open Web Index could secure this critical information infrastructure and restore Europe’s informational sovereignty, as well as ‘have a stimulating impact on digital innovations, in the field of search engines and for the European start-up and internet economy’.

These are just a few examples of possible parts of a public ICT ecosystem. Based on truly sustainable, data-protection-friendly busi-

ness models and green IT, it would serve citizens, companies and the state hand-in-hand. Such an infrastructure could rapidly scale up globally, with its inherent interoperability and data portability, if it is well done. It could provide a different environment for trustworthy and responsible AI services in favour of the common good—in favour, that is, of vulnerable people on a vulnerable planet.

Reinhard Messerschmidt is an interdisciplinary social scientist holding a doctorate in philosophy. At the science-policy interface, he works on topics at the intersection of ethics, technology assessment, research and innovation, and sustainability.

Stefan Ullrich is group lead of the research group @jwi_riot at the Weizenbaum Institute for Networked Society. As an informatician with a minor degree in philosophy, he critically examines the impact of ubiquitous information-technology systems on society.

THIRTEEN
MACHINE LEARNING SHOULD INCREASE
HUMAN POSSIBILITIES

INTERVIEW WITH ELENA ESPOSITO

Sociologist Elena Esposito suggests shifting the focus of artificial intelligence to machines as communication partners. Interview by Florian Butollo.

Butollo: *Artificial intelligence is said to deliver answers on questions such as the right levels of taxation, reasonable urban planning, the management of companies and the assessment of job candidates. Are the abilities of AI to predict and judge better than those of humans? Does the availability of huge amounts of data mean that the world becomes more predictable?*

Esposito: Algorithms can process incomparably more data and perform certain tasks more accurately and reliably than human beings. This is a great advantage that we must keep in mind also when we highlight their limits, which are there and are fundamental. The most obvious is the tendency of algorithms, which learn from available data, to predict the future by projecting forward the structures of the present—including biases and imbalances.

This also produces problems like overfitting, which arises when the system is overly adapted to the examples from the past and loses the ability to capture the empirical variety of the world. For example, it learned so well to interact with the right-handed users it has been trained with that it does not recognise a left-handed person as a possible user.

Algorithms also suffer a specific blindness, especially with regard to the circularity by which predictions affect the future they are aimed to forecast. In many cases the future predicted by the models does not come about, not because they are wrong but precisely because they are right and are followed.

Think, for example, of traffic flow forecasts in the summer for the so-called smart departures: black, red, yellow days, etc. The models predict that on July 31st at noon there will be traffic jams on highways, while at 2 am one will travel better. If we follow the forecasts, which are reliable and well done, we will all be queuing up on the highway at 2 am—contradicting the prediction.

This circularity affects all forecasting models: if you follow the forecast you risk falsifying it. It is difficult to predict surprises and relying too much on algorithmic forms risks limiting the space of invention and the openness of the future.

Do you see political dangers in relying too much on AI? Is the current hype around the subject a sign of the loss of our sovereignty as societies?

The political dangers are there, but they are not determined directly by technology. The possibilities offered by algorithms can lead to very different political outcomes and risks—from the hype about personalisation promising to unfold the autonomy of individual users to the Chinese ‘social credit’ system, which goes in the opposite direction.

What are your recommendations for using AI in the right way? What should policy-makers consider when formulating ethical guidelines, norms and regulations with this in mind?

Heinz von Foerster had as ethical imperative ‘Act always so as to increase the number of possibilities’. Today more than ever it seems to me a fundamental principle. Especially when we are dealing with very complex conditions, I think it is better to try to learn continuously from current developments than to pretend to know where you want to go.

And incidentally, machine-learning algorithms also work in this way. In these advanced-programming techniques algorithms learn from experience and in a way programme themselves—going in directions that the designers themselves often could not predict.

What is a reasonable expectation of AI? What can we hope for and how can we get there?

What I expect with respect to AI is that the very idea to artificially reproduce human intelligence will be abandoned. The most recent algorithms that use machine learning and big data do not work at all like human intelligence and do not even try to emulate it—and precisely for this reason they are able to perform with great effectiveness tasks that until now were reserved for human intelligence.

Through big data, algorithms ‘feed’ on the differences generated (consciously or unconsciously) by individuals and their behaviour to produce new, surprising and potentially instructive information. Algorithmic processes start from the intelligence of users to operate competently as communication partners, with no need to be intelligent themselves.

Elena Esposito is professor of sociology at the University of Bielefeld and the University of Bologna. Her current research on algorithmic prediction is supported by a five-year advanced grant from the European Research Council.

