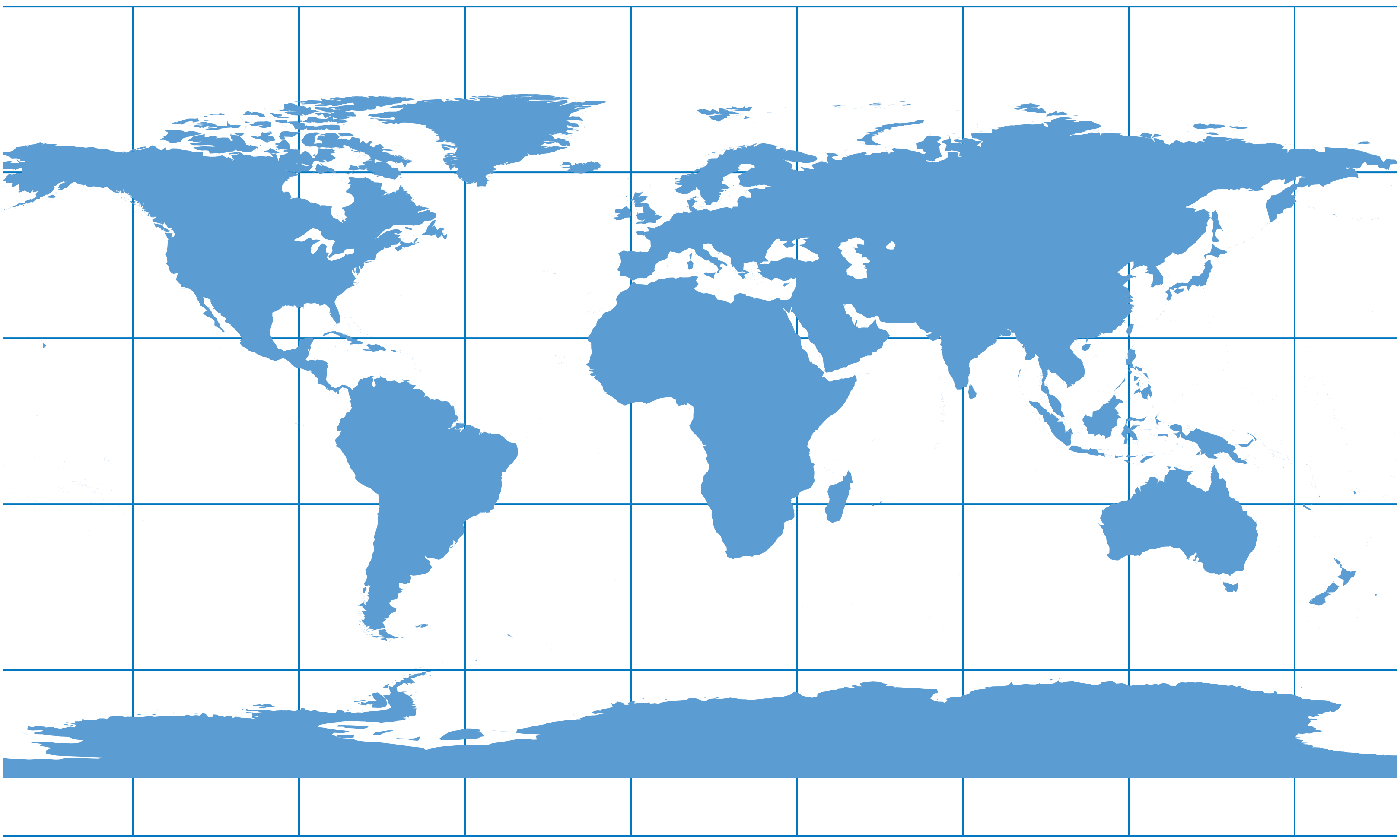
E/2018/50/Rev.1
ST/ESA/370

Department of Economic and Social Affairs

World Economic and Social Survey 2018

Frontier technologies for sustainable development



United Nations
New York, 2018

Chapter II

Managing the promises of frontier technologies

Introduction

Emerging technologies—as introduced in chapter I—will have a profound impact on sustainable development. The present chapter assesses the impact of a few emerging technologies within the contexts of specific countries where they have been developed and deployed in recent years. While these technologies are redefining work, promoting prosperity, improving environmental sustainability and transforming social interactions, they are also presenting equity-related and ethical challenges and are likely to have large direct and spillover effects on the rest of the world as well.

New technologies possess immense potential for achieving sustainable development, but…

New technologies and products possess immense potential but inevitably bring with them uncertainties, risks and unintended and unanticipated consequences. For example, while the discovery and use of fossil fuels have led to a revolution in transportation and a manifold increase in the speed of travel, those fuels have also contributed to global warming and climate change. Similarly, plastics have transformed storage but remain a major source of pollution; processed sugar has expanded our dietary options but its use has led to a higher prevalence of obesity; and the use of fertilizer has resulted in an improvement in crop yields but also in the contamination of rivers and oceans. Indeed, technologies, as these examples illustrate, are seldom neutral—they solve certain problems but create others.

…they also generate uncertainties, risks and unintended consequences

This chapter identifies the opportunities and challenges associated with the advances produced by a few frontier technologies (but whose selection in no way diminishes the importance of other such technologies). The focus is on a specific product, namely, the electric vehicle (EV), and automation and digital technologies, which encompass suites of applications. The aim is to illustrate their promise, their impact in terms of economic and social trade-offs, and their potential spillover effects across sectors and countries.

This chapter identifies the potentials and challenges associated with EVs, automation and digital technologies

While EVs offer a viable alternative to conventional cars with respect to reducing CO_2 and other greenhouse gases emissions, this technology is not yet a viable alternative to conventional vehicles in terms of price and convenience. EVs are unlikely to contribute significantly to reducing CO_2 emissions without a more dramatic shift away from the fossil fuels that they use to recharge their batteries. The promise of a frontier technology may therefore remain unfulfilled without the requisite shift from fossil fuel to renewable energy technologies.

Automation promises to increase labour productivity, income and prosperity. However, its actual and potential impacts on labour demand, production of goods and services, concentration of market power, and wealth and income distribution raise economic and social concerns. The level of apprehension regarding the future of work appears to be growing particularly high in industries and economies where the speed and spread of technological changes have been rapid. Anxieties in this regard are driving social and political discontent, as manifested in many advanced economies, with significant spillover effects on the rest of the world.

Also explored in this chapter is the promise embodied by digital technologies and artificial intelligence (AI) — possibly the final frontier of new technologies — which possess the immense potential to minimize human errors and biases in decision-making processes. On the other hand, AI-powered social platforms are also used to produce targeted advertisements which manipulate human behaviours or spread misinformation, so as to undermine social cohesion, peace and stability. While an automated decision system can improve the efficiency of public agencies, it also runs the risk of reinforcing existing biases and exclusion.

Technological change is inevitable, but humanity can still become better prepared to manage its benefits and risks

The solution, however, is not to steer clear of or to stifle technological progress. Technological change is inevitable, but humanity can become better prepared to manage this inevitability. For example, there need to be more concerted efforts to reduce our dependence on fossil fuels and transform EVs into a more substantive tool for achieving environmental sustainability. Appropriate investments in human capital — through which to foster the acquisition of new skills and knowledge not just by a few privileged groups but by all — can help enable societies to create new jobs and embrace automation without unwarranted fear. Strengthening institutions and mechanisms that play a key role in the determination of wages and benefits — including labour unions, collective bargaining processes and labour regulations (e.g., minimum wage legislation) — can help ensure a more equitable and balanced distribution of gains between labour (employees) and capital (employers), translating into wage growth and robust social protection. Complementary investments are also needed to redefine property rights or even create new ones for the various forms of digital content (Vickers and Ziebarth, 2017) and to develop new ethical, legal and regulatory frameworks for managing algorithms, machine learning and artificial intelligence.

Actions taken by the Governments of the technological frontier countries shape global standards for managing frontier technologies

While the opportunities and challenges discussed in this chapter are related mainly to technologies developed and used in countries at the technological frontier, they are increasingly becoming universal opportunities and challenges, affecting all of humanity. The chapter underscores the special role of Governments in those frontier countries, which are leading innovation in technologies that will affect people and prosperity — indeed, the entire planet. Governments in the handful of developed countries and the few large developing countries that are innovating new technologies and rolling out their applications for businesses and consumers will need to encourage and incentivize innovations that are critical for humanity, while minimizing their unintended adverse economic and social effects. The actions taken by those Governments will shape global standards for managing frontier technologies. The countries that, while not leading innovation, are adopting and using those technologies, will be able to learn from the successes as well as the failures of the frontier countries and in turn pursue national policymaking on appropriate and complementary investments in institutions, regulations and standards so as to achieve their own development objectives.

Electric vehicles: panacea or target of misplaced hope?

Massive consumption of fossil fuels has improved the quality of our lives, but it has also led to global warming and climate change

The twentieth century witnessed a massive increase in the consumption of fossil fuels to power transportation, machines and electricity generation. In essence, fossil fuel now underpins human existence. But while improving the quality of life, it has also emerged as the largest contributor to greenhouse gas (GHG) emissions, global warming and climate change — an adverse (and unanticipated) consequence of meeting the needs of the modern world. Although this adverse climatic impact has become more evident, efficiency- and profit-related considerations are perpetuating the dependence on fossil fuel in the absence of viable alternatives.

There is a growing recognition that humanity must reduce its dependence on fossil fuel to achieve environmental sustainability. Increasingly, many developed countries, and a few large developing countries such as Brazil, China and India, are taking concrete steps to reduce CO_2 emission and enhance environmental sustainability, in line with their commitments to the Paris Agreement[1] adopted under the United Nations Framework Convention on Climate Change.[2] The quest for environmental sustainability is driving technological breakthroughs in (a) energy efficiency and conservation practices, (b) carbon-free or reduced-carbon energy resources and (c) carbon capture, either from fossil fuels or from the atmosphere, and carbon storage.

Electric vehicles (EVs) hold out hope for reducing GHG emissions

Electric vehicles — using carbon-free or reduced-carbon energy resources — represent a technological breakthrough, and a possible small step towards achieving the larger goal of environmental sustainability.[3] EVs replace internal combustion engines with battery-powered or battery-assisted engines which emit significantly fewer or no tailpipe greenhouse gas emissions. Requiring less or no fuel combustion and relying for the most part on electricity, EVs also boost energy efficiency for road transport and thereby contribute to the attainment of a wide range of transport policy goals, such as national energy security, especially for countries that import fossil fuels; and noise reduction and improved air quality, particularly in large cities. Given that the transport industry accounts for 23 per cent of global energy-related GHG emissions (International Energy Agency (IEA), 2016), policy support for EVs, which is growing, represents the kind of urgent action to combat climate change envisaged under Sustainable Development Goal (SDG) 13. EVs may also contribute indirectly to reducing inequality among countries (as envisaged under SDG 10) by reducing the high costs of climate change imposed on those that are both climate-vulnerable and low-income.

While EVs hold out hope for reducing CO_2 and other GHG emissions at the consumer level, their widespread use may not necessarily lead to a significant reduction in those emissions — especially if the breakthrough they represent remains an isolated phenomenon. Use of these vehicles might lead to a reduction in CO_2 emissions in large cities, improving air quality and yielding other benefits to urban residents; but if their batteries continue to be recharged with fossil fuel-generated electricity, overall emission levels could very well remain largely unchanged. For EVs to deliver on their full potential to reduce emissions and generate environmental sustainability, a fundamental shift to renewable energy sources will be needed.

1 See Adoption of the Paris Agreement in United Nations Framework Convention on Climate Change (2015).

2 United Nations, *Treaty Series*, vol. 1771, No. 30822.

3 EVs include battery, plug-in hybrid, range-extended and fuel cell electric vehicles.

EVs are making inroads

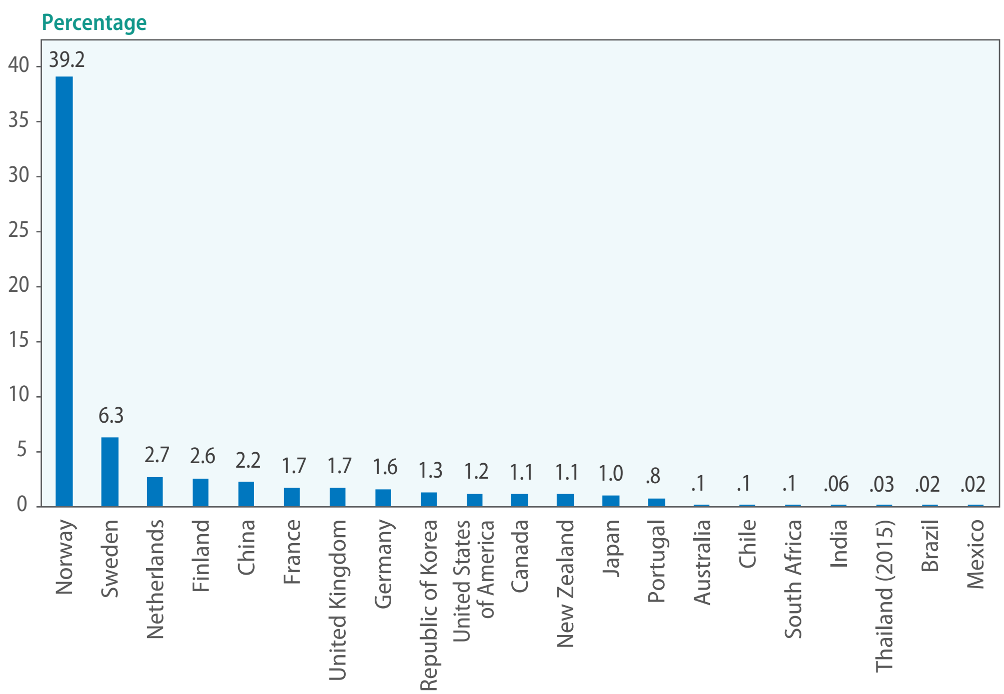
It is hard to imagine that the comforts and conveniences of modern life could have emerged in the absence of the automobile. However, the significant expansion of the use of automobiles during the twentieth century occurred without much consideration being given to their impacts on the Earth's atmosphere. Indeed, large quantities of GHG emissions arising from the use of automobiles (as well as from other uses of fossil fuels) have contributed to significant levels of air pollution, global warming and climate change.

At the same time, the adverse impacts of climate change are distributed unevenly across countries and across population groups within countries. While the poorest people, communities and countries are disproportionally affected by climate change-induced extreme weather events, such as tropical cyclones, and long and frequent drought spells (United Nations, 2016b), it is the developed countries and some large developing countries that have been largely responsible for GHG emissions. This reflects what is often referred to as the inequality of climate change (Lowrey, 2013).

In 2017, about 1.1 million EVs were sold worldwide, with the global stock surpassing 3 million

About 1.1 million electric cars were sold worldwide in 2017, with the global stock surpassing 3 million vehicles. China had the largest electric car stock: constituting about 40 per cent of the global total (IEA, 2018b, p. 9). With a nearly 40 per cent market share in 2017, Norway achieved the highest market penetration of EVs in the world (see figure II.1). Further, the number of public charging stations in the world has continued to increase: from 322,000 in 2016 to 430,000 in 2017.[4]

Figure II.1
Market share of electric cars (battery electric and plug-in hybrid), selected countries, 2017



Source: UN/DESA, based on IEA (2018b), table A.7.

4 Of the 430,000 stations, 320,000 were slow charging and 110,000 were fast-charging (IEA, 2018b).

With improving technologies, EVs will become more affordable and the stock is forecast to reach 125 million–200 million by 2030

A wider variety of EVs — with high-energy batteries and improved operating software — are now available at more affordable prices. Based on increasing public and private expenditures on research and development (R&D) on EV-related technologies — particularly on safer batteries with higher energy density—EVs are expected to become more attractive, with an increase in the variety of vehicle types and sizes in the near future, and, again, at affordable prices. The stock of light EVs is forecast to reach 125–220 million by 2030, but this will largely depend on the level of policy ambition directed towards achieving climate goals (IEA, 2018b, p. 11). UN/DESA predicts that the market share of EVs in the new car market reaches between 5 and 17 per cent in 2025 amid a high level of uncertainty regarding battery capacities, public support measures, oil prices and public acceptance of EVs as the prime mode of transportation, as well as the total demand for all types of vehicles.

EVs may not reduce economy-wide emissions

Replacing conventional cars with EVs will not automatically lead to CO_2 emissions reduction

Replacing internal combustion engine cars with EVs will not lead automatically to the reduction in CO_2 and other GHG emissions at national and global levels, for that reduction depends not on tailpipe emissions of vehicles but rather on so-called total life-cycle emissions and electricity generation structures at national or local levels.

What matters is the life-cycle emissions, not the tailpipe emissions, of vehicles

The total life-cycle emissions (also known as well-to-wheel emissions) of an EV include the emissions derived from its manufacture, battery production, operation, maintenance and disposal, and all of its energy consumption. Lang and others (2013) estimate that, from the perspective of the life cycle, the fuels usage phase (i.e., operation) accounts for most of the total energy consumption of a single EV, followed by the fuels production and transportation stage. That is, the key determinant for reducing GHG emissions is the share of electricity obtained by EVs from renewable sources, including hydro sources, to charge their batteries.

Hawkins, Gausen and Strømman (2012) estimate that the life-cycle global warming potential (GWP)[5] of EVs whose batteries are powered by coal electricity falls somewhere between that of small and large fossil fuel-driven vehicles. On the other hand, the GWP of EVs powered by natural gas or low-carbon energy sources is lower than that of the most efficient internal combustion engine vehicles. A related study focusing on three regions of China (Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta) found that the benefits of switching to EVs are maximized in regions with high proportions of hydropower generation and that, where the proportions of hydropower are nil or small, the per-kilometre consumption of EVs are lowest over their life cycles when the batteries were recharged from natural gas-fired sources, compared with coal-, oil-, biomass- and garbage-fired sources (see Lang and others, 2013). A more recent study in which the total life-cycle emissions of EVs were evaluated found that, even where high volumes of coal are used, EVs produce one quarter fewer emissions than diesels.

5 The global warming potential (GWP) has been developed to allow comparisons of the global warming impacts of different gases. It is a measure of how much energy the emission of one ton of a gas will absorb over a given period of time, relative to the emissions of one ton of CO_2. The GWP of CO_2 is set at 1, regardless of the time span used. See United States Environmental Protection Agency, "Greenhouse gas emissions: understanding global warming potentials", available at www.epa.gov/ghgemissions/understanding-global-warming-potentials.

CO_2 emissions reduction is maximized when EVs use renewable energy sources to recharge their batteries

These studies indicate that the reduction of CO_2 emissions through replacement of internal combustion engine cars with EVs is maximized when those EVs use only renewable energy sources to recharge their batteries (see table II.1). In other words, it is the structure of electricity generation that is the most important determinant of the capacity of EVs to produce positive environmental impacts. Thus, the deeper penetration of EVs into the auto market will not lead automatically to significant reductions of GHG emissions.

Table II.1
Total lifecycle and tailpipe emissions: Internal combustion engine vehicles and EVs

	ICEVs	EVs
Tailpipe emissions	Yes	No
CO_2 tailpipe emissions (grams/kilometre)[a]	255	0
Estimated CO_2 reduction if there is an all-car switch from internal combustion engine vehicles to EVs in the United States (percentage)	16.2	
Life-cycle emissions if batteries are charged with electricity produced from coal	Quantity lies between the quantities for small and large internal combustion engine vehicles	
Life-cycle emissions if batteries are charged with electricity produced from natural gas and renewable resources (including hydropower)	Fewer emissions than from the most efficient internal combustion engine vehicles	

Source: UN/DESA, based on national and international sources.
a For the average-sized passenger car in the United States of America.

Whether EVs contribute to a meaningful reduction in carbon footprints will also depend on how quickly consumers accept EVs as their preferred mode of transportation. In the rural areas of countries covering a large territory—e.g., Australia, Canada and the United States of America—EVs are not yet viable options for many users owing to (a) the relatively short distance that can be travelled by a EV with a single charge and (b) the unavailability of EV charging stations in remote areas. Because of their high prices, EVs are currently affordable only by affluent households but the demand for EVs even among the affluent is highly sensitive to factors such as tax incentives and subsidies. EVs have yet to become a financially viable option for middle-income households. Furthermore, the remarkable growth in shale oil and gas production could halt the future growth of oil prices, making internal combustion engine cars more attractive choices to the user. With the United States accounting for 80 per cent of the increase in global oil supply to 2025, users "are not yet ready to say goodbye to the era of oil" (UBS Limited, 2017).

Policies for making EVs a viable alternative

Policies play a key role in making EVs a reasonable alternative to conventional cars

Policies and incentives have played a key role in making EVs a reasonable alternative to conventional vehicles. Since 2010, Governments in both developing and developed countries have been offering potential EV buyers various incentives. The financial incentives include zero or lower taxes. The non-financial incentives include exemptions from access restrictions to urban areas, dedicated parking opportunities, and preferential access to bus lanes and high-occupancy vehicle lanes.[6] Public R&D expenditures on EVs will continue to play a critical role. The International Energy Agency (IEA) has identified Brazil, Canada, China, India, Japan, the Republic of Korea, South Africa, the United States of America and 20 European countries as having implemented at least one of these incentives to popularize EVs.

Quantitative targets are encouraging EV production and deployment

Policies include setting up long-term targets for EV production, sales or imports, and mandates and regulations

Several Governments announced medium- to long-term targets for EV production, sales or imports, as well as mandates and regulations aimed at achieving those targets. In July 2017, the Government of France announced that it would end the sale of petrol and diesel vehicles by 2040. In October 2017, the city of Paris announced its plan to ban all petrol and diesel cars from Paris by 2030, underscoring that large cities like Paris will need speedier phase-outs of cars with internal combustion engines because of rising levels of nitrogen oxides, a major risk to public health. At the same time, the United Kingdom of Great Britain and Northern Ireland announced its plan to ban all new sales of petrol and diesel cars (including all types of hybrid cars) and vans from 2040. Almost every car and van on the road will need to produce zero emissions by 2050 (United Kingdom, 2017a; 2017b).[7]

In September 2017, the Government of China announced that it is developing a long-term plan to phase out vehicles powered by fossil fuels, but without setting a timeline for a ban. It is considering a dual-credit scheme for manufacturers for their production of more fuel-efficient gasoline cars and new energy vehicles — EVs, including plug-in hybrid and fuel cell models. The scheme is complex and is undergoing changes, but automakers whose annual production is over 50,000 will be assessed as regards new energy vehicle production (International Council on Clean Transportation, 2016; IEA, 2018b, pp. 23–25). The 2020 target translates into about a 4–5 per cent market share in annual car sales.

At least 12 countries have set EV deployment targets and introduced mandates and regulations

At least 12 other countries — including Austria, Denmark, Germany, India, Ireland, Japan, the Netherlands, Norway, Portugal, the Republic of Korea, Spain and the United States — have set EV deployment targets as part of their clean energy and mobility plans. In the United States, 10 States have set their own targets, although there are no national targets. The cumulative assessment of these targets (as developed by the Electric Vehicles Initiative (EVI)), if achieved, suggests the deployment of 13 million EVs in these countries by 2020 (IEA, 2017, EV support policies annex and p. 23).[8]

In 2009, the EVI was established at the intergovernmental level under the Clean Energy Ministerial. As at May 2017, the Initiative had 10 member Governments: Canada, China, France, Germany, Japan, the Netherlands, Norway, Sweden, the United Kingdom and the United States. In 2017, it launched the EV30@30 campaign, which set the collective goal for all EVI member countries of achieving a 30 per cent market share for EVs in the total of all passenger cars, light commercial vehicles, buses and trucks by 2030. IEA is coordinating this important initiative (IEA, 2017, pp. 9–10).

Despite these initiatives, EVs have still a long way to go before they can exert a significant impact on global CO_2 emissions. The global stock of EVs is estimated to have accounted for only about 0.2 per cent of the total number of passenger cars and light trucks

6 Both financial and non-financial incentives change from year to year and are too numerous to list. For details, see Thiel, Krause and Dilara (2015).

7 It should be noted that in 2011, the United Kingdom became the first country to announce its intention that fossil fuel car and van sales should end by 2040.

8 The figure should be considered as only tentative, since national plans of these countries are frequently revised and other countries may join this group of countries.

in 2016; hence, policy support will remain critical for further encouragement of wider use of EVs.

Major global auto makers made important announcements on EV deployment in response to government initiatives

In response to these government initiatives, major global original equipment manufacturers also made important announcements on EV deployment targets. The announcement by Volvo — the Swedish auto brand which is currently owned by Zhejiang Geely Holding Group, a Chinese multinational automaker — of its plan to manufacture only EVs and hybrid cars from 2019 onward has been hailed as the beginning of the end of internal combustion engines' dominance of motor transport after more than a century.

Governments will need to further expand tax, monetary and other incentives in order to render EVs more attractive

Given the prevailing high prices of EVs, Governments will need, increasingly, to provide and expand tax and monetary incentives and other benefits in order to make EVs more attractive. The availability of publicly accessible fast chargers is still limited across countries, with drivers of EVs typically preferring home or workplace chargers (IEA, 2018b, chap. 3). Hence, Governments will also be required to encourage the private sector to improve the charging infrastructure, so that the recharging of EV batteries becomes as convenient and rapid as the refuelling of tanks at conventional gas stations. Furthermore, there is the need for increasing R&D investments in EV research to render them commercially viable for middle-income households without the support of tax subsidies. More importantly, Governments in both developed and developing countries must intensify efforts to shift their energy source from fossil fuels to renewables in order to establish EVs as an important innovation within the context of reducing GHG emissions and improving environmental sustainability.

Is automation a double-edged sword — promoting growing prosperity while fostering growing inequality?

AI, machine learning and robotics have expanded automation to new areas of work

Advances in the field of artificial intelligence (AI) offer the opportunity to expand automation to new areas of work, which has the immense potential to generate productivity gains and economic growth. Such technological advances can also have beneficial impacts on working conditions and health by sparing human labour from having to carry out physically and psychologically demanding tasks.

Historically, automation has created winners and losers, necessitating adequate policies and institutions

The history of automation suggests that technological change typically generates a trade-off between efficiency and equity. That trade-off, which creates winners and losers, requires adequate policies and institutions to minimize the impact on those adversely affected. On this depends the achievement of the Sustainable Development Goals (SDGs), particularly the promotion of sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all (Goal 8) and reduction of inequalities (Goal 10).

However, the potential to automate certain tasks — and possibly entire occupations — does not signify a commensurate disappearance in the aggregate number of jobs, since the automation process will also create new tasks and offer productivity gains, which in turn will spur additional demand for labour. These new employment opportunities could offset the number of jobs lost to automation. On the other hand, the adjustment will not be guaranteed without the aforementioned adequate policies, and may also turn out to be too slow and painful for workers adversely affected by automation. Moreover, the types of new jobs created, especially in the service sector, may not be as well paid as, or of similar quality to, those eliminated by automation.

Ongoing automation could represent a continuation of the automation process that began in the late 1980s with the onset of computerization and a more intensive use of robots. The challenges to average workers and to income distribution that automation may pose therefore need to be considered in the context of the challenges and long-term trends observed since the 1980s.

AI, machine learning and robotics, in expanding to new domains, offer huge opportunities for automating work processes both manual and cognitive (Brynjolfsson and McAfee, 2011; 2017).[9] Routine tasks entail predictable procedures framed by specific rules and are therefore easy to automate; and the automation process which began (as noted above) with computerization in the late 1980s had largely been confined to routine tasks (Autor and Dorn, 2013). However, the use of industrial robots, which accelerated during the 1990s and 2000s in advanced economies (Graetz and Michaels, 2015), has led to the disappearance of many routine tasks in the automotive, electronics and metal product manufacturing industries, which were traditionally performed by low- and medium-skilled workers. Computerization and automation are no longer confined to routine jobs (Brynjolfsson and McAfee, 2011). Deep learning algorithms, for instance, can now outperform humans in detecting patterns in big data. In a new automation age, engagement in cognitive tasks will no longer be the exclusive prerogative of humans.

A debate persists on the extent to which jobs could be automated and replaced by machines

An intense debate persists on the extent to which jobs could be automated and replaced by machines (Bruckner, LaFleur and Pitterle, 2017) — a debate that is centred on analysing potential impacts on specific tasks versus entire occupations. The analytical results for tasks are different from those for occupations and both sets of results are not always comparable. The estimated impacts on the tasks tend to be lower than those estimated for entire occupations. Notwithstanding these differences, the new wave of automation will extend to many non-routine tasks, putting low and medium skills more at risk than higher ones.

Many technological, economic and social factors determine the extent and pace of automation

Many factors will determine the extent and pace of automation and its impact on tasks and occupations (McKinsey Global Institute, 2017a), including technical feasibility, advancements of AI in speech and sensory perception, the cost of automation, wage and labour-market flexibilities, potential productivity gains, and improvements in quality and convenience of automation, as well as regulatory frameworks and behavioural factors. There will be a need to differentiate between the labour saving and labour augmenting

9 Artificial intelligence (AI), one of the most significant and potentially disruptive technological developments observed in recent years, encompasses technologies as diverse as "intelligent" stock trading, human speech recognition and self-driving vehicles. At the core of current AI applications, is machine learning—where machines become capable of learn from large amounts of data—and whose development has been very rapid. This has been made possible through the evolution of the Internet and the increase in the availability of large amounts of digital data for analysis. In machine learning, rather than learn from human beings, machines utilize all available information, to achieve the ability to perform a wide range of activities. Often cited in this regard are the advantages machine learning offers in the health sector. A computer can make use of 600,000 medical reports or millions of patient records for pattern recognition and compare the results with a specific case to determine the best treatment plan. In the financial industry, automation is also being taken seriously, as both an opportunity and a threat. Analysts, for example, are becoming redundant, given that new algorithms—often performed by small start-up companies—have the potential to automate a large part of their work. Decisions regarding loans are now being made by software which can take into account a wide variety of detailed data on a borrower, instead of simply using the classic credit score.

effects of automation: While labour saving automation may increase unemployment, labour augmenting technologies may increase the demand for high-skilled workers, leading to the polarization of labour markets and an increase in wage inequality. In less developed economies, where the levels of both wages and adoption of frontier technologies are low, automation will likely take root at a slower pace (World Bank, 2016).

Slow adjustment in labour markets

Automation displaces labour in some tasks, but creates demand for labour to perform non-automated tasks

Automation will require adjustments in labour markets, and adjustment costs may be particularly onerous for less skilled workers. A new wave of automation, which will cause displacement of labour in some tasks, is expected to result in a reduction of wages and, ultimately, of the share of labour in national income. However, this initial displacement effect can be offset by productivity gains (Acemoglu and Restrepo, 2018a), entailing a reduction of production costs and an increasing demand, as the economy expands, for labour to perform non-automated tasks (either in the same sector or other sectors), including those requiring adaptability, common sense and creativity (Autor, 2015). During the early stages of the computerization process, for example, the task composition in the United States reflected a shift towards more interpersonal and communication-intensive activities (Michaels, Rauch and Redding, 2013). This computerization process, which started in the 1980s, generated demand for analytical and interactive work, as routine tasks became automated (Autor, Levy and Murnane, 2003).

The impact of automation on employment growth will vary across sectors. The increased use of information and communications technologies (ICT) in the manufacturing sector, for example, is associated with slower growth in manufacturing jobs, but greater use of ICT in the service sector has had little or no negative impact on employment growth. However, concerns remain regarding whether automation will create a sufficient number of new jobs to compensate for the jobs lost owing to automation. When computer spreadsheets began to replace manual bookkeeping, and the bookkeepers were replaced by data-processing staff as well as software and hardware professionals in the same sector or industry, the impact on the total number of jobs was minimal. Ongoing AI-led automation, especially automation of non-routine tasks, is likely to have a larger negative impact, on both routine and non-routine jobs.

A study shows that each high-tech job creates 4.9 additional jobs in other occupations…

The silver lining is the potential spillover effects of automation on other sectors, generating additional demand for goods and services (Acemoglu and Restrepo, 2018a). There are estimates suggesting that each high-tech job creates 4.9 additional jobs in other occupations (Moretti, 2010, as cited in Berger and Frey, 2016) which explains partly why recent job growth in countries that are members of the Organization for Economic Cooperation and Development (OECD) has been largely concentrated in non-technology sectors (Berger and Frey, 2016).

The pace of adoption of emerging technologies will determine the time required for labour-market adjustments. Adjustments to new tasks will require new skills, which the workforce may lack, especially when technology requires higher skills and when the educational system cannot anticipate future demand for skills. The mismatch between skills and new tasks not only slows down employment and wages adjustments, but can also undermine potential productivity gains (Acemoglu and Restrepo, 2018a).

…but automation is likely to affect low-skilled, low-wage workers, exacerbating income inequality

Even when automation leads to higher productivity and increased demand for goods and services from non-automated sectors, aggregate demand in an economy may still stagnate or even fall. Automation is likely to adversely affect low-skilled, low-wage workers, who tend to have a higher propensity to consume than high-skilled, high-income workers. A permanent decline in labour income of low-skilled workers may therefore depress economic growth. Automation in developed countries may also reduce imports from low-income countries which rely on relatively low-cost labour, displacing workers in their export sectors and potentially exacerbating income inequality among countries.

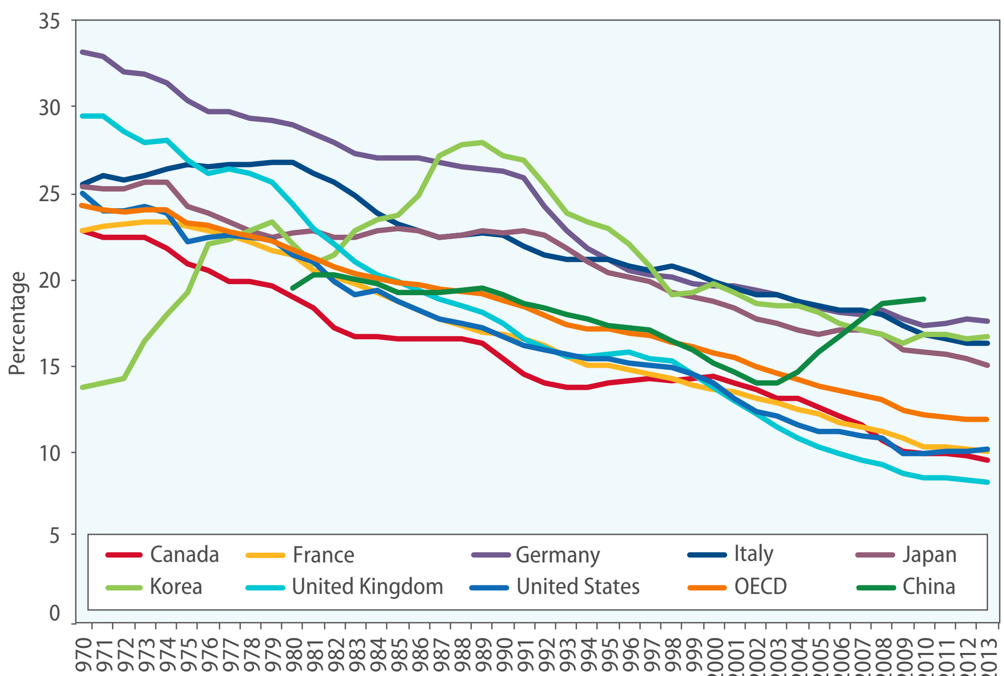
Automation and the future of work

The empirical evidence since the 1980s has illustrated how automation has led to a reduction in jobs in routine intensive occupations and to the polarization of labour markets, which contributed to a significant increase in wage inequalities. Computerization and robotization in the 1980s and 1990s reduced the demand for labour that performed routine tasks (Autor, Levy and Murnane, 2003). Industrial robots, introduced in late 1980s, automated many of the labour-intensive tasks in manufacturing, including machining, welding, painting, palletizing, assembly, material handling and quality control (Graetz and Michaels, 2015). This led to a long-term secular decline in the share of labour in routine-intensive occupations. For instance, in OECD countries, the share of employment in the manufacturing sector decreased from about 25 per cent in the 1970s to about 10 per cent in 2013 (OECD, 2015) (see figure II.2). While various factors contributed to the decline in manufacturing jobs in OECD countries, automation is considered a key underlying factor (OECD, 2012).

Parallel to the elimination of routine tasks, job growth in OECD countries slowed down in the medium-skill category over the past 20 years (see figure II.3). On the other hand, job opportunities increased at both ends of the skill spectrum in those countries (OECD, 2017e), suggesting increased polarization of skills.

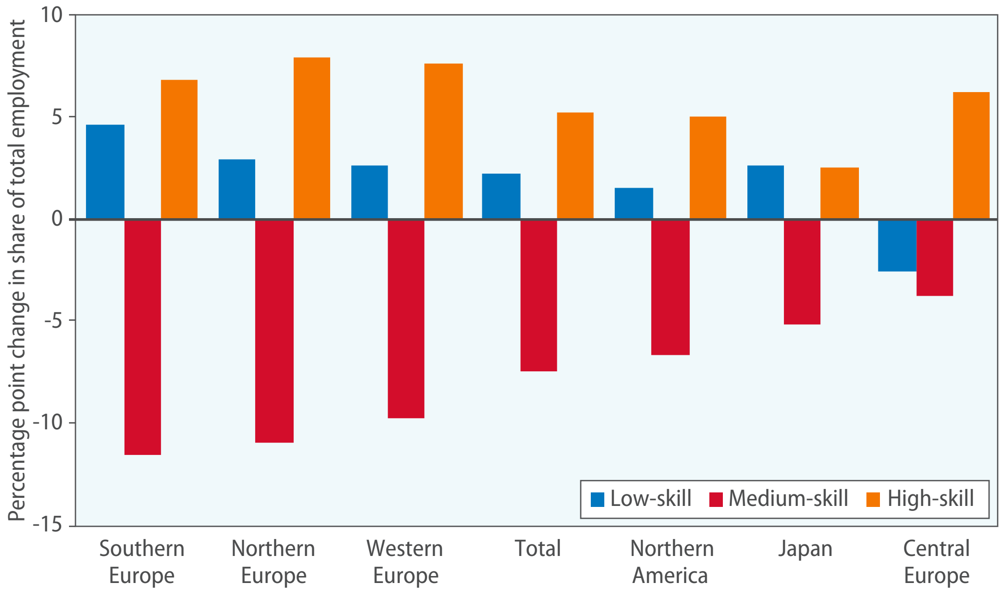
The introduction of automation slowed down job growth in the medium-skill job category

Figure II.2
Decline in manufacturing jobs in OECD countries and China, 1970–2013



Source: OECD (2015).

Figure II.3
Polarization of labour markets in OECD countries, 1995–2015



Source: OECD (2017e).

At the same time, the number of jobs increased in both the high- and low-skill categories

This polarization of skills can be explained largely by computerization which has favoured relatively higher-skilled labour. On the other hand, the increase in incomes for higher-skilled workers led to an increased demand for goods and services in other sectors, performed largely by low-skilled labour. There was a surge in low-skill service jobs as well as in work involving manual non-routine tasks, which were not susceptible to computerization. This hollowing out of the middle of the wage distribution is well documented for the United States (Autor, Katz and Kearney, 2006; Acemoglu and Autor, 2011) and for European countries (Goos, Manning and Salomons, 2014). Recent work by the World Bank (2016) indicates that labour markets have also become polarized in many developing countries since the mid-1990s, with the share of medium-skill occupations declining (Bruckner, LaFleur and Pitterle, 2017).

Skill-biased technological change and wage inequalities

Job polarization has widened wage gaps among workers with different levels of education and…

The polarization of skills has widened the wage gaps between workers with a college education and those with a high school education in the United States and other developed countries since the 1980s. While workers with a high school degree earned about three quarters of the wages of their college-educated counterparts in 1980, the former now earn only about half as much. The trend holds true for other OECD countries, although there are considerable cross-country differences in respect of the skill premium. Since 1970, the real wages of high-skilled workers have risen faster not only than the wages of medium-skilled workers, but than those of low-skilled workers as well. In the majority of developed countries, wage inequality (as measured by the 90:10 ratio) is higher today than 40 years ago, with the bulk of the increase having occurred in the 1980s and 1990s. In the United States, where wage inequality is significantly higher than in any other developed economy, the 90:10 ratio rose from 3.65 in 1979 to 5.05 in 2016, owing mainly to higher wage increases at the top of the distribution (Bruckner, LaFleur and Pitterle, 2017). On the other hand, there is growing evidence that differences in labour-market institutions, as reflected, for example, in union density, employment protection and minimum wage laws, play an important role in containing wage gaps and skill premiums (Koeniger, Leonardi and Nuziata, 2007; OECD, 2017e; Bruckner, LaFleur and Pitterle, 2017).

With further automation, the polarization of labour markets is expected to continue, which would potentially further aggravate wage inequalities. In this regard, the International Federation of Robotics estimates that the number of robots in advanced economies could increase fourfold by 2025. Should the spread of robots be as rapid as anticipated by several analysts, the negative consequences for the aggregate employment and wage will be significantly stronger than those that have been observed so far (Acemoglu and Restrepo, 2017).

…polarization is expected to continue in the future, further aggravating wage inequalities

Declining share of labour income

An automation process can further increase income and wealth inequalities through its effect on the distribution of income between capital and labour. Automation is inherently capital-intensive. Increased capital intensity in production of goods and services typically increases the total return on capital and the share of capital income in gross domestic product (GDP). As discussed above, AI, machine learning and robots are expected to lead to a substitution of labour by capital for certain skills, with direct and severe consequences for income distribution.

Automation has also reduced the labour share of national income in developed countries since the 1990s

The labour share of income was stable in developed economies until the 1980s, despite important variations in the short and medium terms. Since then, however, the share of labour income has been declining consistently across advanced economies for several decades (OECD, 2012; IMF, 2017), contradicting the notion of a stable labour share of income in the long term. For example, between 1990 and 2009, the labour share of national income declined in 26 out of 30 advanced countries for which data were available (see figure II.4). During that period, the median (adjusted) labour share of national income across these countries fell from 66.1 to 61.7 per cent (OECD, 2012).

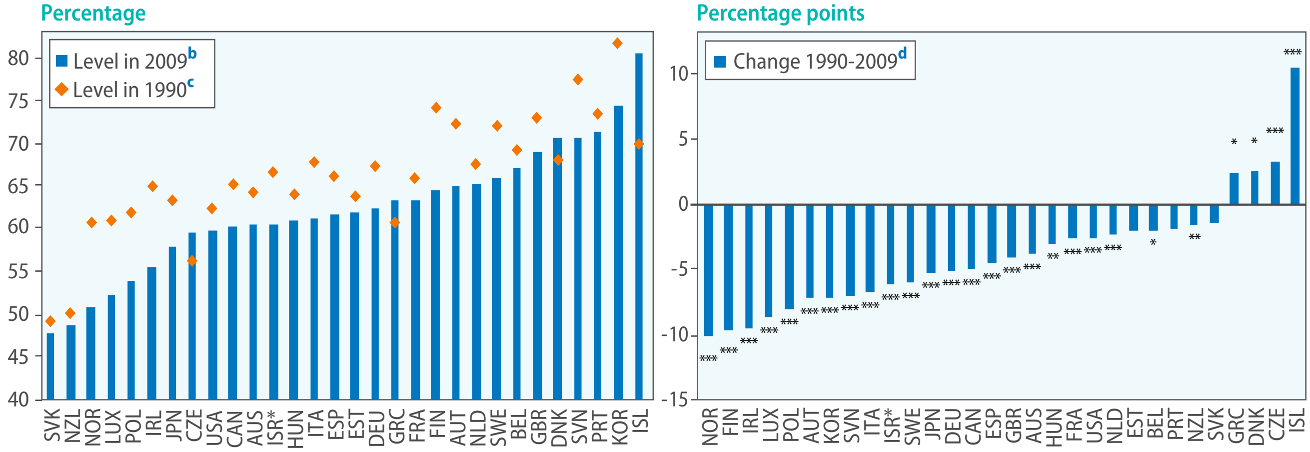
In some emerging and developing economies, the decline in the labour share of national income is even more pronounced than in advanced economies, with considerable declines in Asia and Northern Africa (ILO, 2011). In a recent study, Karabarbounis and Neiman (2013) found that labour share in GDP had declined in 42 out of 59 countries, including China, India and Mexico, and concluded that, as advances in information technology reduced the cost of plants, machinery and equipment, firms became more capital-intensive and reduced the number of employees. A high degree of substitution between capital and labour — particularly less-skilled labour (Brynjolfsson and McAfee, 2011) — explains the declining share of labour income.

The labour share also declined in some developing countries, particularly in Asia and Northern Africa

Further, there is evidence that wage and productivity growth diverged during the same period in most advanced economies. In the majority of G20 countries for which data are available, the aggregate growth of real wages was significantly slower than that of aggregate productivity (see figure II.5), even taking into account the dynamics of relative prices, which thus accounted for the decline in the labour share (ILO and OECD, 2015). However, the divergence between productivity growth and wage growth is presumably less pronounced for high-skilled workers because of the skill premium received by those workers. This notwithstanding, increasing productivity is not a sufficient condition for an increase in the real wages of the average worker.

Since 2000, wage and productivity growth have diverged, particularly for lower-skilled workers

Figure II.4
Decline of labour share in OECD countries, 1990–2009[a]



Source: OECD (2012), figure 3.1.
Notes:

a Graphs represent three-year averages, starting and ending with indicated years.

b Germany and Iceland: 1991; Estonia: 1993; Poland: 1994; Czech Republic, Greece, Hungary, Slovak Republic and Slovenia: 1995; Israel: 2000.

c Portugal: 2005; Canada and New Zealand: 2006; Australia, Belgium, Ireland, Norway and Sweden: 2007; France, Iceland, Israel, Poland and the United Kingdom: 2008.

d ***,** and * indicate significance at the 1 per cent, 5 per cent and 10 per cent levels, respectively. Statistical significance refers to the coefficient of the time trend in a bivariate regression on annual data with the labour share as dependent variable. The wage of the self-employed is imputed assuming that their annual wage is the same as for the average employee of the whole economy.

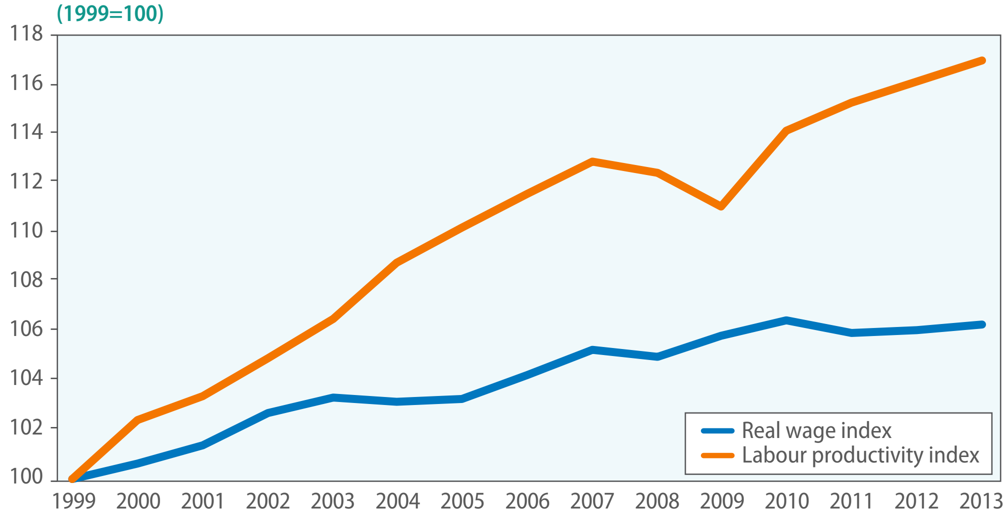
Declining labour-force participation

Automation is also partly associated with the declining labour-force participation observed in OECD countries

Declining labour-force participation—that is, decline in the proportion of people who are employed or looking for work—across advanced economies is a concerning labour-market trend. The declining labour-force participation rate has been associated partly with the automation process, entailing a painful adjustment by workers whose skills became redundant. In OECD countries, labour-force participation has trended downward, particularly for members of the prime male labour force between the ages of 25 and 64. This decline became more pronounced starting in the 1990s and, later, with the unfolding of the global financial crisis.

Reductions in the demand for labour, especially for lower-skilled men, appear to be another critical component of the decline in prime-age male labour-force participation in developed economies. This is consistent with the observation that technological change weakened demand for less-skilled labour, principally in the manufacturing sector, making job polarization a major contributor to the declining labour-force participation rates (Krause and Sawhill, 2017). There is also evidence that prime-age men often choose not to work under a given set of labour-market conditions. For example, supply-side factors, such as increased participation in social programmes (e.g., those offering disability insurance or food stamps) and the setting of a high reservation wage, contribute to low labour-force participation in the United States (ibid.).

Figure II.5
Average wages and labour productivity in selected G20 countries, 1999–2013



Source: ILO and OECD (2015).
Note: Data refer to Australia, Canada, France, Germany, Italy, Japan, Republic of Korea, the United Kingdom of Great Britain and Northern Ireland and the United States of America. Real wage growth is calculated as a weighted average of year-on-year growth in real average monthly wages in the advanced G20 economies. For a description of the methodology, see ILO (2015), appendix I.

Policies for protecting employment and wages

Policymakers need to minimize the adverse impact of automation

AI, machine learning and the new age of robotics present a number of policy challenges to minimizing their potential negative impact on employment, wages and inequalities. Political reaction to frontier technologies can, in theory, slow down or even prevent their adoption and development if they do not promote shared prosperity (Acemoglu and Restrepo, 2018a). It is therefore important to focus on policies that have the potential to minimize the impact of these emerging technologies on employment and income distribution.

Build forward-looking and inclusive education systems

The education system should be more inclusive, particularly for disadvantaged groups, including youth

Automation will require a constant upgrading of workforce skills. However, many workers whose jobs are partially or fully automatable lack the skills and ability they would need to switch to the higher-skill jobs created by automation. There are considerable cross-country variations in the impact of automation on wage inequalities since the 1980s, which is partly explained by differences in terms of the availability of educated and skilled workers. Given that access to higher and better education is often determined by the socioeconomic background of parents, the educational system needs to be more inclusive in the age of automation so as to ensure that socioeconomically disadvantaged population groups have opportunities to acquire the skill sets that are relevant in markets for increasingly automated jobs.

As unemployment and the risk of falling below a poverty line are particularly high for youth, whose education and career choices have yet to be made, younger generations need to be made aware of the potential of automation, including the changes that it will generate in the labour market. In this regard, Governments could contribute to developing an educational system that facilitates the acquisition of basic skills and capabilities which are adaptable and less at risk of becoming automatable, thereby helping youth identify the skills that will be complementary to the automation process (Brynjolfsson and McAfee, 2014).

Private firms and workers have roles to play in reducing the persisting skills mismatch

Governments could implement policies that encourage private firms and workers to invest in continuous learning and development of skills in areas where demand remains unmet. For example, shortage of data scientists and business translators has been a concern in many economies (McKinsey Global Institute, 2017b). Digital skills will become essential for a majority of workers; and forward-looking educational policies, and appropriate governmental support, should aim at reducing the persisting skills mismatch, particularly in sectors adversely affected by automation. In the age of AI-driven automation, non-automable skills demanding social and emotional intelligence as well as creativity, will become increasingly important. While they do not always require higher levels of educational attainment, greater investments in such skills will be required within conventional educational systems, which do not always value this type of intelligence and creativity.

Expand social protection coverage

A robust and effective social protection system should be in place

A robust and effective social protection system can help minimize both the adverse impact of technological changes on specific income groups and the resistance to those changes (Korinek and Stiglitz, 2017). During the adjustment process, medium- and low-skilled workers typically face longer periods of unemployment, at least until their skills are upgraded. Social insurance programmes can be critical in providing affected workers with sustenance during these transitional periods involving joblessness. In addition, there will be a clear need for coverage by targeted social protection schemes of specific sectors and locations where the displacement effect is stronger. Active labour-market policies—including, e.g., job placement services, special labour-market programmes and wage subsidies—can help facilitate adaptation during the transition period, especially of less-skilled workers. Automation may create jobs in one region, while eliminating them in other regions; and if the workers affected are to avail themselves of new job opportunities, they may be faced with high search and relocation costs. They would therefore require help in relocating to those areas where employment growth is faster. Policies to facilitate such geographical mobility could target housing and moving costs, among others (Berger and Frey, 2016).

Social protection schemes in developed countries will need to evolve if they are to confront the new realities of non-standard employment conditions. In many advanced economies, work under temporary contracts, part-time jobs and self-employment often do not qualify for full social protection. As frontier technologies increasingly favour self-employment, part-time work and new types of employment based on sharing-economy models, there will be a need for the extension of social protection coverage, funded with tax revenues, to workers whose conditions of employment are non-standard.

The universal basic income, which would provide a regular unconditional cash grant to every individual, has gained fresh importance in this new age of automation, given the risks it presents of loss of employment and decline in wages. The empirical evidence needed to assess the impact of such an initiative is still lacking. In addition, Governments will need to increase tax revenues substantially in order to cover all of the population with a UBI high enough to tackle poverty.

Build stronger labour-market institutions

Stronger labour-market institutions should be (re)built

Empirical evidence suggests that labour unions can play a major role in ensuring a fair distribution of national income (OECD, 2012). They give workers, in particular less-skilled workers, a stronger bargaining power in negotiating wages that match their productivity and ensure minimum labour standards. Since the 1970s, the density of labour unions has decreased considerably in developed economies; and de-unionization, along with automation, has been considered an important factor in the decline in the wages of lower-skilled workers (Acemoglu, 2000). As AI and other similar technologies are expected to exert pressure on wages, especially on the wages of medium- and low-skilled labour, workers will need more effective representation.

Introduce progressive and innovative taxes

Income tax schemes should be more progressive towards the very top of the income distribution

Reducing income inequalities will also require more progressive income tax schemes. Disposable income has increased much faster at the top of the income distribution, owing not only to skill- and capital-biased technological change, but also to the less progressive taxation that was introduced during the past three decades. Income tax schemes are needed to become more progressive, especially towards the very top of the income distribution. By reducing the accumulation of capital and wealth of top income-earners—and consequently the return on accumulated wealth-tax schemes of this type reduce not just post-tax income inequality, but future pre-tax income inequality as well.

Taxing robots and taxing capital returns from patent monopolies to reduce income inequality have been suggested

The concept of taxing *robots* has gained traction (Guerreiro, Rebelo and Teles, 2018), as suggested in a 31 May 2016 report of the Committee of Legal Affairs to the European Parliament. In that report, the Committee introduced a motion for a European Parliament resolution in which the Parliament would emphasize that "consideration should be given to the possible need to introduce corporate reporting requirements on the extent and proportion of the contribution of robotics and AI to the economic results of a company for the purpose of taxation and social security contributions". Taxes of this type could generate the resources required to retrain workers and expand employment in the health-care and education sectors. While a number of entrepreneurs have supported such a requirement and while some countries have taken concrete steps in this direction, developing a common understanding of the definition of "robot" remains a challenge. If such a definition is not clearly established, a tax on robots may simply induce their producers to bundle the components of this new technology with other types of machinery. It has also been suggested that a robot registry be created to keep an accounting of the loss of jobs performed by humans and facilitate compensation for the loss of revenues through a supplementary tax, which could be levied at the corporate or the robot level. As indicated above, the speed at which automation is being introduced poses a challenge. Hence, slowing down automation and creating tax disincentives to counter technology's displacement effect on employment could be sensible policy options and serve as the basis for a policy that is suitable for some countries. However, the effect might be only temporary, inasmuch as countries will need to keep pace with technological development if they are to compete in international markets.

The returns on capital earned by innovators are an important source of income inequality (Korinek and Stiglitz, 2017). Taxing return on capital—especially excess return earned from patent monopolies—may be more conducive to a balancing of income distribution. A suggestion in the same vein has been to shorten the term of patents, which would accelerate the entry of innovations into the public domain and their accessibility and limit monopolistic income advantages.

Digital technologies: a Pandora's box?

Social media and online platforms, backed by AI, may have opened a Pandora's box of ethical issues

AI, powered by algorithms and machine learning, is defining the future of digital technologies, with economic and social activities increasingly being shifted from the physical world to the digital space. Computer codes and algorithms are the key drivers of various applications of the technologies in that space—ranging from activities on social media and other online platforms to automated decision systems used in public agencies. While advances in digital technologies offer great benefits in terms of efficiency and information sharing, they may have also opened up a Pandora's box of ethical issues related to fairness and inclusion, privacy and autonomy, and accountability and transparency.

Addressing ethical issues is crucial to ensuring social inclusion and cohesion, and political stability

Deployment of the algorithms driving social media and other online platforms may lead to discrimination against specific groups of people and an undermining of informed decision-making. Addressing these concerns is crucial to ensuring social inclusion, social cohesion and political stability, as envisaged under Sustainable Development Goals 16 and 17. The increasing use of automated decision systems in the areas of job recruitment and criminal justice also runs the risks of further reinforcing biases against minority groups and exacerbating social inequalities. Fulfilling the imperatives of leaving no one behind—the cardinal objective of the 2030 Agenda—and of reducing inequality, promoting social inclusion and eliminating discriminatory practices, as envisaged under SDG 10, requires urgent action to address these challenges.

Rapid advances in various digital technologies, increasingly underpinned by artificial intelligence, render existing regulatory frameworks, social norms and ethical standards inadequate. Societies must develop new ethical standards on the use of those technologies; and policymakers and the public must reflect concretely on the meaning of fairness and accountability as they will apply in digital space. While ethical and social norms vary across countries, the new standards should be grounded in internationally agreed instruments, such as the Charter of the United Nations and the Universal Declaration of Human Rights,[10] which provide the framework for the protection of and respect for human rights.

The need to address the undesirable impacts of digital technologies, as already manifested on social media platforms and in algorithmic decision systems—and to avert even more negative consequences as applications of those technologies proliferate more widely—puts every society at a critical juncture. Through an examination of the issues at hand, the present section charts a way forward towards enabling a society to benefit from the efficiency gains to be achieved through digital technology while addressing collateral equity and ethical challenges. There is a clear need for policymakers to step in and for public debate to determine the appropriate balance among efficiency, equity and ethics.

Online platforms: connecting or disconnecting people?

Social media have changed the way we interact with each other

Social media and other online platforms have greatly changed the way social interactions and the spread of information are carried out. It is increasingly obvious that the use of social media and other online platforms—which have remained unregulated for years—can have negative societal consequences. Those platforms have facilitated the spread of misinformation and hate speech and created so-called echo chambers which have contributed to the polarization of society and have possibly influenced elections. Further,

10 General Assembly resolution 217 A (III).

they have collected massive amounts of data which are used by the platforms themselves, by advertisers and by other third parties, with ramifications extending to privacy, freedom and, potentially, to the very foundations of democracy. While disruptions of this type are not a new phenomenon, the ease of communication on social media and the ability to deploy big data-driven algorithms to sway or rouse large population groups have caught policymakers by surprise.

Echo chambers and the spread of misinformation

The Internet, while creating a global village, is also increasingly fostering the formation of isolated digital communities through use of algorithms to shape social media interactions (El-Bermawy, 2016). People form these "islands" to interact with other people who possess and share similar views. This dynamic, which locks participants into personalized feedback loops or the above-mentioned echo chambers, has arguably widened societal divides, allowing different groups to live in their own cognitive bubbles and reinforcing confirmation biases. Those algorithms also have an enormous impact on how the information consumed by people is selected. Algorithms take advantage of human vulnerabilities: they can manipulate a user by presenting content that may either reinforce or contradict that user's opinions.

Echo chambers have been cited as one of the key contributors to the political polarization experienced by many developed countries in recent years. Recent studies have revealed how Facebook users come to inhabit highly polarized closed communities (Quattrociocchi, Scala and Sunstein, 2016) and how people who communicate on Twitter become disproportionately exposed to the tweets of like-minded users (Halberstam and Knight, 2014). This can propel people towards ever more extreme viewpoints, a tendency referred to as algorithmic radicalization and also as enclave extremism (Sunstein, 2007). Such online platforms are well suited to the amplification of the voices of a small group, a process in which algorithms play a key role. For example, when a user engages with a certain type of content, the algorithm-based recommendation system will pull that user towards more extreme or more radical content (Nicas, 2018).

Echo chambers have also contributed to the undermining of objective expertise and the spreading of misinformation (OECD, 2017c). These platforms favour content that grabs the user's attentions and maximizes engagement, regardless of its accuracy, and whatever users see in their newsfeeds has been algorithmically curated. The algorithms, combined with automated accounts (so-called bots), ensure that false information spreads fast (Vosoughi, Roy and Aral, 2018).

Social media have disseminated hate speech and content that incites violence

These issues are being confronted not only in developed countries but in many developing countries as well. Hate speech, content that incites violence, and disinformation targeting specific minority groups have been disseminated rapidly on social media in several developing countries, often with devastating consequences.[11] False information is disseminated differently in developing as compared with developed countries. This is due to limited availability of official information for fact-checking and the lack of public confidence in news media sources. Language barriers, higher illiteracy rates and the relative higher cost of securing Internet access serve to limit the amount of time people can devote to obtaining truthful information (World Wide Web Foundation, 2017).

11 See, for example, the statement by the Chairman of the Independent International Fact-Finding Mission on Myanmar, Marzuki Darusman, at the thirty-seventh session of the Human Rights Council on 12 March 2018.

It is important that the spread of misinformation and hate speech on social media platforms be addressed, while at the same time respecting freedom of speech and avoiding undue censorship. The clear need to ensure accountability for content and to apply content moderation should be balanced by an awareness of the dangers of surveillance, censorship and suppression of free speech.

Targeting advertisements, discrimination and manipulation

Digital technology firms generate revenue by selling ads based on personal information

The largest digital technology firms generate revenue by selling advertisements based on personal information collected on online platforms or by search engines. The reliance on advertising as the primary business model for revenue generation creates adverse incentives for online platforms, which are often faced with a trade-off between protecting user privacy and generating ad revenue. While users have benefited from free access to these platforms, they bear the hidden cost of ceding control of their personal data.

The consent agreements governing the operation of these data exchanges are often opaque and their terms are consequently unclear to users. Consumers have come to trust companies with vast amounts of data of a highly intimate nature, which can result in the loss of ownership of those data. It is particularly difficult for users to anticipate the ways in which the personal information that is extracted might be used and reused by third parties. Intense data collection can enable advertisers to increase consumer satisfaction by targeting relevant advertisements to specific user groups. However, targeted advertisements also raise many ethical issues, with implications for consumers related to privacy, manipulation and potential discrimination (Plane and others, 2017).

Advertisers can target specific groups of people to view, or to be excluded from viewing, their ads

Their ability to identify specific users has made it possible for advertisers to target specific groups of people to view—or to be excluded from viewing—their ads. Not only is this practice questionable from an ethical perspective but it sometimes runs counter to certain civil rights laws (Angwin and Paris, 2016). ProPublica has demonstrated that it is indeed possible for advertisers to exclude certain categories of users when placing a housing advertisement on Facebook, which may constitute a violation of United States federal legislation, namely, the Fair Housing Act (Angwin, Tobin and Varner, 2017). Both Facebook and Google subsequently disallowed advertisers the use of characteristics such as ethnic "affinity" as a means of preventing ads related to housing, employment or financial services from being seen. However, Speicher and others (2018), investigating the different targeting methods offered by Facebook, have shown that even without relying on sensitive attributes, an advertiser can still create highly discriminatory ads.

Targeted ads can have implications for democratic processes and elections

The potential of targeted advertisements can have implications as well for democratic processes and elections. The massive amounts of data derived from social media platforms have enabled researchers to build accurate psychological profiles of individuals, which enable personalized political advertising. This entails tailoring messages to the specific interests and vulnerabilities of particular voters in order to manipulate them, invade their privacy and undermine their agency, autonomy and freedom. Personalization algorithms of this type must strike an ethical balance between coercion and support for the decision-making autonomy of users (Lewis and Westlund, 2015).

Those who own and control this kind of information and data wield real power over people. The accumulation of personal data by credit agencies, social media companies and other entities has significant implications with respect to who has the right to own and monetize personal data. Even if, technically speaking, people are the owners of their personal data, they may not be able to exercise control over those data, and this has important implications. A key means of preserving the ability of people to exercise that control is to ensure that, for example, they have the right to data portability, and hence the ability to transfer their data from one service provider to another.

An opaque surveillance infrastructure enables platforms to exercise power over people

The data collection that facilitates targeted advertising is underpinned by an opaque surveillance infrastructure, which enables platforms to exercise immense power over individuals and, potentially, over the whole of society. To the extent that people are unaware of their rights and the options available for protecting their privacy, they are understandably surprised when confronted by the magnitude of the data concerning them that are available on those platforms (LaFleur, Iversen and Jensen, 2018). Data security and protection of privacy are factors critical to ensuring that social media and other online platforms can be trusted and held accountable. Lack of data protection has, in several instances, compromised the personal information of users. The lack of adequately enforced contractual restrictions on third-party users of data is an issue that must be addressed.

Automated decision systems: addressing human bias or reinforcing it?

Automated decision systems are now being used for decision-making in public and private domains

Automated decision systems, based to varying degrees on AI, are being used increasingly for decision-making in many domains. In the private sector, automated systems are being deployed to facilitate hiring practices, and in the provision of loans. Public sector automated systems contribute to decision-making in the criminal justice system, the education sector and the system of social and children's protection services. While in some cases automated decision systems have improved efficiency, consistency and fairness, in others, they have reinforced historical discrimination and obscured undesirable behaviour (Rieke, Bogen and Robinson, 2018).

Replacing human judgment with machines: issues of efficiency, explainability and bias

Automated decision systems can improve efficiency by enabling firms and public institutions to make more informed decisions in a shorter period of time. Indeed, Brynjolfsson and McAfee (2017) argue that in conducting various tasks, machines outperform humans in minimizing bias and error. They contend that while people should remain in the loop for the purpose of common-sense checking, most decision-making should be assigned to algorithms. The belief in the superiority of machines over human judgment is shared by Kahneman (2011), who argues that the decision-making process of humans is "noisy". Especially when the amount of information is large and it is costly for humans to process that information, algorithms will outperform humans. Kahnemann therefore argues that humans should be replaced by algorithms "whenever possible".[12]

Automated decision systems can reinforce existing human bias

Proponents of automated decision systems claim that they not only increase efficiency, but also reduce human bias. However, there are many counter-examples which demonstrate how machine learning reinforces existing bias, discrimination and prejudice, and leads to further social exclusion. Data can be biased, as they are often incomplete, skewed or drawn from non-representative samples, and algorithm developers can encode the bias, consciously or unconsciously, when programming the machine learning processes (Campolo and

12 Remarks by Nobel laureate Daniel Kahneman made at the National Bureau of Economic Reseaech inaugural conference on the Economics of AI, held in Toronto in 2017.

others, 2017). The harms inflicted by such bias can be categorized as either (a) harms of *allocation*, arising when a system allocates a certain opportunity or resource to, or withholds it from, a specific group or (b) harms of *representation*, arising when, through technology, the subordination of some social and cultural groups becomes entrenched (see box II.1).

While an individual can be held accountable for a decision, there is no mechanism for ensuring the transparency and accountability of opaque, "black-box" automated decision systems. Machine learning has created a fundamentally different approach to programming (discussed in more detail in chapter I). While this approach has increased programming efficiency, it has also contributed to greater opaqueness. According to Brynjolfsson and McAfee (2017), "machine learning systems often have low interpretability, meaning that humans have difficulty figuring out how the systems reached their decisions".

Greater explainability is required, given the difficulty of deconstructing how an automated system has reached its decisions

In consequence, there is an increasingly loud call for explainability with respect to automated decision systems. However, people in the technology field fear that requiring this technology to be explainable will only slow down progress, reducing the potential of machine learning to address important challenges, such as diagnosing diseases (Weinberger, 2018). To fully tap the potential of machine learning, it is necessary to relinquish the need to understand the systems involved, as it is often literally impossible to explain their operation to the human mind. In this sense, there is a clear trade-off between progress in machine learning as measured by accuracy and efficiency, and the need for explanations and transparency.

Automated decision systems in public agencies

Automated decision-making processes in the public sector can disproportionally affect disadvantaged people

Automated decision systems have radically changed decision-making processes in many public agencies. However, as those systems are being used in high-stakes domains, issues of bias and discrimination have advanced to the forefront of the public debate. Not only are there inherent biases in the data and algorithms used, but automated decision systems are more often deployed in domains of society where they will affect disadvantaged people. According to Eubanks (2018), many of these systems are first tested on low-income households where there is less of an expectation of respect for privacy. Moreover, the increased prevalence of algorithms in the decision-making processes of public agencies can lead to a decrease in their visibility and, at the same time, an amplification of their effects through layering.

Two applications of machine learning in the criminal justice system—namely, as tools for risk assessment and for predictive policing—have been heavily debated. The United States criminal justice system uses a machine learning tool to calculate what is referred to as a risk score, which is then considered by judges in making pretrial, parole and sentencing decisions. In analysing the efficacy of this tool, Angwin and others (2016) found that the predictions were racially biased and that the predictions made by the system affected black and white defendants differently. While the data used by the software do not include an individual's race, there are other elements of the data that correlate to race, which leads to racial disparities in terms of predictions. Predictive policing provides another powerful example of how algorithms can amplify historical bias. Using machine learning techniques, police departments try to predict the locations of future crimes. Historically, crime data are biased against certain minorities. As a result, the algorithms driving this type of program—which entails learning from previous crime reports—are sometimes trapped within a vicious feedback loop, which results in the over-policing of certain neighbourhoods (Lum and Isaac, 2016).

Box II.1

Two types of potential harm arising from automated decision systems

Harms of allocation — inflicted when a system allocates a certain opportunity or resource to, or withholds it from, a specific group — are well known within the context of automated decision systems. For example, banks using automated systems to evaluate mortgage applications have ended up unfairly denying mortgages to certain minorities or people from a specific geographical area (Harney, 2008).

Recently, more attention has been given to problems related to harms of representation including social stigmatization, where technology reinforces the subordination of some social and cultural groups. In recent examples of such harm, an image recognition programme labelled the faces of several black people as belonging to gorillas; and in a Google Images search for "CEO", the first woman to appear was Barbie! While these "errors" were quickly fixed by the companies and characterized as simple glitches within the systems, they highlight a deeper problem associated with bias in automated systems. Noble (2018) has explored, in particular, how negative stereotypes of black women are codified in search engine algorithms.

In many cases, representational harm can have allocative consequences. For example, the perpetuation of stereotypes regarding a certain group can reduce the employability of the members of that group. Use of automated decision systems in public agencies poses this risk, as the historical data often reinforce past representational harms, which generates economic or identity-based impacts (Reisman and others, 2018).

Table II.1.1

Potential harms arising from algorithmic decision-making

	Example	Impact
	Harms of allocation	
Credit discrimination	Withholding specific credit offered to members of certain groups	**Economic loss and loss of opportunity**
Employment discrimination	Filtering candidates by geographical proximity, leading to exclusion of minorities	
Insurance and social benefits discrimination	Increasing auto-insurance prices for workers on a night shift	
Housing discrimination	Housing advertisement displayed only to certain groups	
Education discrimination	Ads for only for-profit colleges presented to low-income individuals	
	Harms of representation	
Confirmation bias	Image search results for "CEO" consist only of male images	**Social stigmatization**
Increased surveillance	Use of predictive policing which results in the presence of more police in minority neighbourhoods	
Stereotype reinforcement	Word-embedding models reveal gender stereotypes	
Dignitary harms	Emotional distress arising from bias or from a decision based on incorrect data	

Source: UN/DESA, based on Future of Privacy Forum (2017).

Automation must guarantee key values of society, such as fairness, justice and due process

While it is reasonable, in some cases, for public institutions to employ automated decision systems, and private and public information utilized in the systems, to increase efficiency, it is important to understand that both data and algorithms encode bias. As a result, minorities and vulnerable groups can end up being affected disproportionately; and as long as these systems are not built to explicitly dismantle structural inequalities, they are more likely to intensify those inequalities dramatically (Eubanks, 2018). While some have argued that technological fixes to the bias and explainability challenge posed by automated decision systems are available, those fixes remain largely theoretical. Indeed, addressing bias requires more than a technological fix: it requires an understanding of the underlying structural inequalities. In essence, the use of automated decision systems has outpaced the development of the frameworks required to understand and govern them. Given these concerns, there have been serious calls for a cessation of the use of unaudited black box systems in core public agencies, at least until key values such as fairness, justice and due process are guaranteed (Campolo and others, 2017).

Policies for producing socially responsible digital technologies

While the benefits of digital technologies are significant, it is important for policymakers and other stakeholders to proceed with adequate caution in this domain. Rather than accept decisions made by machines uncritically, society needs to construct the mechanism best suited to combine machine intelligence with human wisdom. There is a tendency of many in the technology industry to highlight the negative consequences of "dumbing down" AI for the purpose of providing transparency. Notwithstanding their concerns, it is imperative that a full understanding of the implications of automated decision-making be achieved, even if this entails a slower pace of progress in the field of AI.

It is important that the debate focused on ethical norms and regulatory architecture be shaped not only by leading technology companies but by public debate and Governments as well. Policymakers have a significant and proactive role to play in developing the legal and ethical frameworks needed to govern the evolution and use of digital technologies.

Make privacy laws fit for the digital age

More extensive privacy protection for consumers and the promotion of a data ownership model are needed

Consumers are in need of more extensive privacy protections. The current system, which relies on individualized informed consent, is problematic, as people often do not understand the privacy-related consequences of providing their data. It is increasingly difficult to perceive those consequences since, through the advances in machine learning, seemingly superficial data can be linked with other data in such a way as to reveal highly private information. A third-party rating agency can help protect privacy by offering consumers the opportunity to better understand the consequences of data sharing and thereby enable them to make more-informed decisions on whether or not to share their data.

There are also calls for the promotion of a data ownership model under which people can share or sell their own data if they so desire. However, this could enable firms to take advantage of a consumer's financial situation to secure access to their personal data. One alternative would be the adoption of a data protection law providing individuals with more fundamental rights regarding the processing of all personal data. This would be crucial to ensuring that data privacy is understood to be a right, not a luxury affordable only by some. A balance must be sought, however, with regard to the ethical responsibility to share data for the common good. By sharing data, people will enable technology to attain longstanding goals for the public good, such as achieving a cure for cancer (Domingos, 2015). Hence, it might be important to strike a balance between respecting the need for privacy and making data available as a public good.

The European Union is at the forefront of the discussion on privacy and data protection

The European Union (EU) is at the forefront of the discussion on privacy and data protection. The General Data Protection Regulation (GDPR),[13] which was agreed by the European Parliament and the Council of the European Union in April 2016 and became enforceable in Europe in May 2018, will require social platforms to change the way that they collect data from their customers and store and deploy them. The predictions are mixed, however, regarding the societal and ethical impacts of the Regulation, as the cost of compliance for companies may be so high as to limit innovation and access to technology within the EU. Some argue that the GDPR will have a negative effect on AI innovation while at the same time failing to protect—or even potentially harming—consumers (Wallace and Castro, 2018). However, the proponents of the Regulation predict that it will provide valuable protections for consumers which will produce a ripple effect extending beyond the EU (Susswein, 2018).

Since all companies with a presence in Europe must implement the rules set out in the Regulation to cover their operation there, it should also be possible for companies to put a system in place for extending the same protection to users elsewhere in the world (as Facebook has hinted that it will strive to do). However, the potential voluntary geographical extension of the GDPR by some technology companies, for the purpose of covering other countries, would not eliminate the need for an international standard on data protection and regulation.

Encourage diversity and ethics education in the technology field

As technology is not value-neutral, it needs to be shaped by diverse communities in order to reduce adverse impacts

There is a disconnect between people who develop technologies and the communities that are affected by those technologies. Since technology is not value-neutral, it needs to be built and shaped by diverse communities so as to minimize adverse social consequences, such as bias, prejudice and discrimination. Indeed, women and minority groups remain underrepresented in the technology field, and policymakers need to be proactive in transcending this status quo.

Some technology industry leaders warn that studying subjects other than science, technology, engineering and mathematics (STEM) would be a mistake for anyone seeking a job within the digital economy. While it is true that advancing artificial intelligence will require greater numbers of people who have digital skills and training in data science, the fact remains that tackling many of the adverse social impacts highlighted above will entail more than just a proficiency in STEM: Not just technical skills will matter, but how one thinks. Critical thinking, cognitive flexibility and creativity will remain important assets in the future. It is also crucial that a greater focus on ethics be incorporated in data and computer science education. There is after all an urgent need in this age of big data for clearer ethical guidelines on research and experimentation that are applicable to both universities and private companies.

13 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Prevent the spread of misinformation and false news

Social media have been under increased scrutiny for their failure to remove misinformation, hate speech and other such content

Social media platforms have come under increased scrutiny for their failure to remove misinformation, illegal content, discriminatory ads and hate speech, as well as for their tolerance of fake accounts. There is a widespread belief within the technology community that artificial intelligence will be the panacea for these technological ills. AI, however, is inadequate for addressing a problem that is so complex and so entangled with its economic, psychological and political roots. Regulation is needed to compel those platforms to take the steps necessary to prevent the dissemination of the fabricated, false and misleading content on their sites. Policymakers in some countries have started to target these issues and make firms accountable for such content. For example, in Germany, under the new Network Enforcement Act ("NetzDG"), which entered into law in June 2017, online platforms face fines of up to 50 million euros if they do not remove "obviously illegal" hate speech and other postings within 24 hours of receiving a notification.

Promote fair, accountable and transparent automated decision systems

Governments and other stakeholders should ensure ethical use of digital technologies

Governments and other stakeholders should apply the foundational principles of fairness, transparency and accountability so as to ensure ethical use of digital technology. Algorithmic fairness is important for ensuring that automated decision-making does not exert discriminatory or unjust impacts across different demographics such as race and gender. Accountability is important for establishing avenues of redress for adverse effects of an algorithmic decision system on individuals or societies. Assigning responsibility, especially in cases of technological redlining, is vital for the rapid redress of discrimination.

Building transparent algorithms capable of explaining their own reasoning can promote transparency. That most automated decision systems are little more than black boxes for the people affected by them is an issue that should be addressed. The basis for the decision-making process taking place within those black boxes should be made comprehensible to those affected; however, many companies have been resistant to laying bare the structure of their algorithms because of commercial sensitivities. New regulations are therefore needed to ensure disclosure. It is also important to enable access for interested third parties to review the behaviour of those algorithms.

Some national and local governments have taken the first step towards creating a framework to govern the public use of AI

Some promising steps have been taken to address these issues. For example, the Article 29 Working Party on the Protection of Individuals with regard to the Processing of Personal Data—an advisory body made up of a representative from the data protection authority of each member State of the European Union, the European Data Protection Supervisor and the European Commission—adopted guidelines on automated decision-making and profiling, including the provision that people should have the right to challenge the decisions and that companies should be able to provide users with an explanation for the decisions reached by automated systems. Along similar lines, the New York City Council passed legislation in December 2017 requiring the creation of a task force to review the use of algorithms by New York City agencies in various public policy decisions and to provide recommendations on how information on agency automated decision systems may be shared with the public. This was the first comprehensive algorithmic accountability bill passed in the United States, and represented an important first step towards creating a framework designed to govern the public use of AI and related digital technologies.