example.pdf

#

TL;DR
Artificial intelligence: How does it work, why does it matter, and what can we do about it? The dynamics of public opposition and acceptance could be important factors shaping AI's long-term development path. The data that drives ML-enabled sectors is often collected by offering users access to services in exchange for data.

#

Executive Summary
- Artificial
intelligence:
How does it
work, why
does it matter,
and what can
we do about it? STUDY
Panel for the Future of Science and Technology
EPRS | European Parliamentary Research Service
Author: Philip Boucher
Scientific Foresight Unit (STOA)
EN
PE 641.547 – June 2020

Artificial intelligence:
How does it work,
why does it matter, and
what can we do about it? Artificial intelligence (AI) is probably the defining technology of the
last decade, and perhaps also the next. The aim of this study
- Artificial intelligence: How does it work, why does it matter, and what can we do about

it? calculus – while balancing a range of considerations about the problem itself and the context of its
solution. First, the engineer needs to find a good way of encoding the problem itself. For the chess-playing
ANN, the engineer needs to express the positions on the board as a signal to be sent to the input
layer. They also need to find a way of interpreting the output as a valid move. This means either
de
- Artificial intelligence: How does it work, why does it matter, and what can we do about it? The dynamics of public opposition and acceptance could be important factors shaping AI's long-
term development path. As with many technologies, public opposition to AI is often explained with
reference to a 'knowledge deficit model', whereby citizens are assumed to oppose technologies
because they do not understand how they work, and their concerns are interpreted as a failure to

appreciate their positiv
- STOA | Panel for the Future of Science and Technology
to SMEs and other firms, to help them to take full advantage of digital opportunities. The AI4EU
consortium was established specifically to support AI development, adoption and collaboration. These initiatives provide a good starting point, and could be complemented by information
campaigns, site visits, business mentoring and other schemes targeting SMEs. • Articulate a development path. For many successful AI firms in Europe, a prominent vi
- STOA | Panel for the Future of Science and Technology
• Assess data quality. Tools can be applied to identify quality issues such as incorrect data labels,
inappropriate biases, illegal material (including information gathered without consent) or 'fake
news'. Depending on the system, material might be removed automatically or flagged for
human review. The difficulty with such tools

is that they are also biased, in this case against data
that does not conform to their definition of quality. Defin

##

Section 1
Artificial intelligence: How does it work, why does it matter, and what can we do about it? STUDY
Panel for the Future of Science and Technology
EPRS | European Parliamentary Research Service
Author: Philip Boucher
Scientific Foresight Unit (STOA)
EN
PE 641.547 – June 2020

Artificial intelligence: How does it work, why does it matter, and what can we do about it? Artificial intelligence (AI) is probably the defining technology of the last decade, and perhaps also the next. The aim of this study is to support meaningful reflection and productive debate about AI by

providing accessible information about the full range of current and speculative techniques and their associated impacts, and setting out a wide range of regulatory, technological and societal measures that could be mobilised in response. AUTHOR

Philip Boucher, Scientific Foresight Unit (STOA),

This study has been drawn up by the Scientific Foresight Unit (STOA), within the Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament.

To contact the publisher, please e-mail stoa@ep.europa.eu

LINGUISTIC VERSION

Original: EN

Manuscript completed in June 2020.

ISBN: 978-92-846-6770-3
doi: 10.2861/44572
QA-01-20-338-EN-N
http://www.europarl.europa.eu/stoa (STOA website)
http://www.eprs.ep.parl.union.eu (intranet)
http://www.europarl.europa.eu/thinktank (internet)
http://epthinktank.eu (blog)

Artificial intelligence: How does it work, why does it matter, and what can wed o about it?

Executive summary

Artificial intelligence (AI) is probably the defining technology of the last decade, and perhaps also
the next. The aim of this study is to support meaningful reflection and productive debate

about AI
by providing accessible information about the full range of current and speculative techniques and
their associated impacts, and setting out a wide range of regulatory, technological and societal
measures that could be mobilised in response. What is artificial intelligence? The study adopts the European Commission's 2018 definition of AI, which is both accessible and typical of contemporary definitions. AI refers to systems that display intelligent behaviour by analysing their environment and
taking action – with some degree of autonomy – to achieve specific goals. Since AI refers to so many techniques and contexts, greater precision is required in order to hold meaningful and constructive debates about it.

##

Section 2

Artificial intelligence: How does it work, why does it matter, and what can we do about it?

very large and complicated, very quickly. Symbolic AI is at its best in constrained environments which
do not change much over time, where the rules are strict and the variables are unambiguous and
quantifiable. Once such example is calculating tax liability. Tax experts and programmers can work
together to develop expert systems that apply the rules that are in force for that tax year. When
presented with data describing taxpayers' income and other relevant circumstances, the tool can
calculate tax liability according to the rules and applying any applicable levies, allowances and
exceptions. 2.1.2 Fuzzy logic: capturing intuitive expertise

In the expert system described above, each variable is either true or false. For it to work, the system
needs to know an absolute answer to

questions such as whether or not the patient has a fever. This could be reduced to a simple calculation of a temperature reading above 37 °C, but reality is not always so clear cut. Fuzzy logic is another approach to expert systems which allow variables to have a 'truth value' that is anywhere between 0 and 1, which captures the extent to which it fits a category. This allows patients to be assigned a rating of how well they fit the category of having fever. The figure might depend on the patient's temperature reading as well as other relevant factors such as their age or the time of day, and it allows the patient to be described as a borderline case. This fuzzy logic is particularly useful for capturing intuitive knowledge, where experts make good decisions in the face of wide-ranging and uncertain variables that interact with each other. They have been used to develop control systems for cameras which automatically adjust their settings to suit

the conditions, and for stock trading applications to establish rules for buying and selling under different market conditions. In each case, the fuzzy system continually assesses dozens of variables, follows rules designed by human experts to adjust truth values and uses them to automatically make decisions. 2.1.3 Good old-fashioned artificial intelligence

Symbolic AI systems require human experts to encode their knowledge in a way the computer can understand. This places significant constraints on their degree of autonomy. While they can perform tasks automatically, they can only do so in the ways in which they are instructed, and they can only be improved by direct human intervention. This makes symbolic AI less effective for complex problems where not only the variables change in real-time, but also the rules. Unfortunately, these are the problems where we need the most

help. Millions of 'if-then-else' rules could not capture all of a doctor's domain knowledge and expertise, nor their continual development over time. Despite these limitations, symbolic AI remains far from obsolete. It is particularly useful in supporting humans working on repetitive problems in well-defined domains including machine control and decision support systems.

##

Section 3

Artificial intelligence: How does it work, why does it matter, and what can we do about it?

calculus – while balancing a range of considerations about the problem itself and the context of its

solution. First, the engineer needs to find a good way of encoding the problem itself. For the chess-playing

ANN, the engineer needs to express the positions on the board as a signal to be sent to the input

layer. They also need to find a way of interpreting the output as a valid move. This means either

designing the output layer so that its signal can always be interpreted as a legitimate move, or

devising a strategy for managing any illegitimate moves suggested by the ANN. If the ML algorithm uses training data, the AI engineer must consider which data to use and how. Where 'data in the wild' is used, they must ensure that it is legal and ethical. Even inadvertent storage

and processing of some content – such as terrorist propaganda and child pornography

– can be illegal. Other data might be subject to copyright, or require 'informed consent' from the owner before it is used for research or other purposes. If the data passes these tests, the engineer must determine whether it is sufficiently large and representative to be suitable for the problem at hand. A dataset for learning to recognise cats should contain lots of pictures from different angles, of different colours, and any labels should be accurate. Finally, the engineer needs to decide how much data to use for training, and how much to set aside for testing. If the training dataset is too small, the ANN can memorise it without learning general rules, so they perform poorly when tested with new data. If the testing dataset is small, there is less scope to evaluate the quality of the algorithm. The AI engineer also needs to make several important decisions about the structure of the ANN and the ML algorithm. The ANN needs enough

neurons and layers to deal with the complexity of the problem. Too few and the ANN will not be able to deal with complex problems, too many and they tend to memorise the training dataset instead of learning general rules. For gradient descent, they need to define how many evaluations to make before deciding on a direction to travel, as well as how far to travel in the chosen direction before re-evaluating. This is known as the 'learning rate'. If it is slower, the algorithm takes more time but makes better choices, like the lost hiker taking small careful steps. If it is faster, it adapts more quickly but might miss important features, rather as though the hiker runs blindly through the fog. The engineer must consider the problem and decide how to balance speed against accuracy. In evolutionary approaches, the AI engineer has to decide the population size and number of games

to play, balancing thorough evaluation against processing burden. They also need to decide how many ANNs to delete per generation, and how combination and mutation are used to create new generations.

##

Section 4

Artificial intelligence: How does it work, why does it matter, and what can we do about it?

Here, it is also worth mentioning that the marriage of AI and robotics is a major area of development for military technologies, notably in autonomous weapons systems. At present, drones are remotely piloted by humans, but this introduces several weaknesses, including communication channels that are vulnerable to detection and attack, as well as much slower human decision and response times than with automated control systems. Full AI command resolves both issues while opening new opportunities such as swarming capabilities which are beyond human capability. Such systems are not beyond today's technical capabilities, but the field is controversial and 'human-in-the-loop' policies dominate, as discussed in the next chapter. 2.3.3 Quantum artificial intelligence

Quantum computers harness the power of

What is quantum computing? simultaneity to quickly find solutions to very
Single bits of data on normal computers complex problems, promising a revolutionary exist in a single state, either 0 or 1. Single increase in computing power. If the problem is to bits in a quantum computer, known as find a one-in-a-trillion combination that works as a 'qubits' can exist in both states at the same
solution, a normal computer would have to check time. If each qubit can simultaneously be
each possibility one by one, while a quantum both 0 and 1, then four qubits together computer can check them all at the same time, in a could simultaneously be in 16 different
single operation. This means they are particularly states (0000, 0001, 0010, etc.). Small
well-suited to problems such as simulating increases to the number of qubits lead to environments, finding solutions, and optimising massive increases (2n) in the number of
them. Since these kinds of problems are

central to AI, simultaneous states. So 50 qubits together
can be in over a trillion different states at developments in quantum computing could enable
the same time. Quantum computing works significant advances in the field. by
harnessing this simultaneity to find
While there have been some promising recent solutions to complex problems very quickly.
breakthroughs in quantum computing, their details
often serve to illustrate how far the technology is from launching on the market. For example, in late
2017, IBM's 50-qubit machine broke industry records by remaining stable for 0.00009 seconds. Two
years later, Google claimed quantum supremacy when their 54-qubit machine completed a
calculation in 200 seconds that might have taken a non-quantum supercomputer up to 10 000 years
to complete. However, while the machine is an impressive proof of concept, it is not yet capable of

performing calculations with specific practical uses. A general-purpose quantum computer would require closer to 1 million qubits operating near absolute zero (- 273 °C). As such, it seems that reliable and useful quantum computers will probably remain unavailable for at least the next decade. Some suggest it is a moving target that will always remain tantalisingly out of reach.

##

Section 5

Artificial intelligence: How does it work, why does it matter, and what can we do about it?

The dynamics of public opposition and acceptance could be important factors shaping AI's long-term development path. As with many technologies, public opposition to AI is often explained with reference to a 'knowledge deficit model', whereby citizens are assumed to oppose technologies because they do not understand how they work, and their concerns are interpreted as a failure to appreciate their positive impacts. Within this model, strategies for achieving acceptance include informing citizens how technologies work while highlighting the benefits they can bring and downplaying the risks. However, these approaches have been criticised as inaccurate and ineffective.16 Inaccurate because public opposition is more often characterised by lack of meaningful

engagement and control than misunderstanding, and ineffective because repeating positive messages without recognising problems can lead to a breakdown of trust and adoption of more entrenched positions. More sophisticated understandings recognise that citizens can adopt more nuanced and active roles than passive 'acceptor' or 'rejecter' of technologies. Public acceptability of AI (and other technologies) is most effectively achieved by engaging citizens early in the development process to ensure that its application is acceptable, rather than developing the technology first and then encouraging citizens to accept it as it is. Similarly, to encourage trust, it is more effective to design safe and secure systems rather than encourage citizens to have confidence in technologies that might let them down later on. How these understandings can inform action in the context of AI will be discussed in the next

chapter. 3.1.4 Identifying fact and fiction

As mentioned in section 2.2.7, ML can be deployed to generate extremely realistic fake videos – as well as audio, text and images – known as 'deepfakes'. The availability of data and algorithms make it increasingly easy and cheap to produce deepfakes, bringing them within reach of individuals with relatively modest skills and resources. The deepfakes themselves are only one side of the problem, as powerful dissemination platforms – also powered by ML in some cases – can spread these materials very quickly. Together, these applications present financial risks, reputational threats and challenges to the decision-making processes of individuals, organisations and wider society.17 The boundaries of the problem are not limited to the fake material itself. Indeed, the very existence of deepfakes introduces doubts about the authenticity of all content, including real

videos. This could raise the bar for evidence – as recordings can be brushed aside as forgeries – and contribute to a broad climate of disbelief and social polarisation. A broader problem can also be identified in differentiating between appearances and reality in the digital age. This includes reliance upon algorithms to gauge and predict performance.

##

Section 6
Artificial intelligence: How does it work, why does it matter, and what can we do about it?
Furthermore, the data that drives ML-enabled sectors is often collected by offering users access to services in exchange for data and exposure to advertisements. As explained in the context of spam detection in section 2.2.6, the more widely a service is used, the more data it can collect and use to improve ML services for users and advertisers alike. These services, in turn, attract more users and the cycle of data collection and service development continues. In this way, market dominance is, in itself, a driver of further market dominance. These market distortions also present a major barrier to users who consider leaving. Unlike consumers who change their internet or electricity provider,
Figure 6 – Global market share by company

with minor cost and temporary inconvenience, those that change their social media provider lose access to the whole network, permanently. This dynamic leads Sources: W3Counter, GSStatCounter, eMarketer. to a rather extreme concentration of resources. At present, social media service providers have access to substantial information about all of their users, significant control over the information that they receive and the choices they have, and even the capability to 'nudge' their emotional states. Users, on the other hand, have limited options and few alternatives to choose from. This dynamic favours incumbents in the market which may be innovative, but can use their dominant position to outcompete or buy-out their competitors, and use their global reach to develop more tax efficient strategies.30

The competitive edge can also be considered on a global scale, which presents some further

challenges as different global actors seek to influence how AI is developed and deployed. That is why global adoption of European values is celebrated (as seen with GDPR), while reliance upon imports is controversial (as seen with 5G infrastructure). In the 'global AI race', the EU is often positioned as struggling for the bronze medal behind the USA and China. Indeed, while Europe does maintain an important role in global AI development, particularly in terms of fundamental research, it is widely recognised that the USA and China dominate the frontline of global AI development. This is often explained in reference to their higher levels of investment, lower levels of data protection, and appetite for application and adoption. However, the 'race' metaphor is limited as it implies a definitive finish line that is the same for all participants, whereas AI is a range of technologies that

actors can deploy in different ways depending on their priorities, contexts and values. Through this lens, it is most important to define the right direction and develop at an appropriate speed. 3.1.10 Distributing costs and benefits The costs and benefits of AI are not always evenly distributed. The previous section showed how network effects can concentrate the benefits of AI in a small number of successful firms.

##

to SMEs and other firms, to help them to take full advantage of digital opportunities. The AI4EU
consortium was established specifically to support AI development, adoption and collaboration. These initiatives provide a good starting point, and could be complemented by information
campaigns, site visits, business mentoring and other schemes targeting SMEs. • Articulate a development path. For many successful AI firms in Europe, a prominent vision of
future success is to be bought-out by a larger firm, often from the USA, as for example with DeepMind which was bought by Google in 2014. European AI could benefit from a reversal of
this trend but, to do so, firms need the resources and confidence to develop, scale-up and
mature. Alongside measures discussed elsewhere to improve access to resources – notably
capital, data and skills – it could help to

articulate an alternative to the buy-out vision. One approach could be to celebrate champions and to assure champions-in-waiting that support will be provided at all stages of development. Individual measures or complete programmes could be mobilised to support maturity, provided by a specialist consortium or elite innovation hub. Targeted public procurement and investment support could also play a role in helping European start-ups to mature and deliver projects with social value. • Adopt an ambitious vision. Incremental developments with minor benefits for certain sectors may distract attention from more ambitious opportunities for greater disruption that could make a serious contribution to grand challenges. For example, if self-driving cars liberate drivers from the steering wheel without substantially disrupting the model, this could represent an immense missed opportunity for implementing a new generation of mobility

services that offer shared door-to-door public transport while reducing the environment, health and mobility burdens associated with privately-owned single-occupant vehicles. • Foster mission-oriented innovation. Achieving such ambitious visions will require not only AI development, but also substantial coordinated effort in multiple sectors and domains, and could be targeted through mission-oriented innovation. These approaches – exemplified by the Apollo programme and contemporary varieties such as the European Organisation for Nuclear Research (CERN) and the Human Brain Project – combine ambitious concrete challenges with elements of 'blue sky' exploratory research. The specific missions should be bold and widely relevant; targeted and measurable; ambitious but realistic; cross disciplinary and sectoral; and include multiple bottom-up solutions. The implementation plan can deploy a wide range

of instruments
–prizes, funding, public procurement and nonfinancial support – and would need to recognise
the specificities of the innovation ecosystem that is relevant to the mission.42
• Create innovation spaces or 'sandboxes'.

##

• Enhance and enforce consumer control. Several policy measures such as the GDPR already empower users with greater control over their data. This could be further enhanced through measures to give users greater control, including transparency about how their data is used to train algorithms and how algorithms are used to process data and make decisions. Consumers of AI products should be protected from false advertising. Further measures could support consumers in exercising these rights, including mechanisms for recourse such as compensation for the misuse of data or products that do not deliver on their promises. 4.1.6 Update mechanisms for ensuring liability
Safety rules are the primary means of protecting consumers from economic or physical damage caused by faulty or dangerous products. When they fail to achieve their task, liability

rules enable consumers to be compensated by the responsible party or an insurance scheme. AI products and services are no different but, for several reasons – including their complexity, opacity, autonomy and learning features – it can be difficult to prove fault and establish liability. The specific challenges of AI may also make it more important to look beyond physical damage to mental and moral damage. Having reliable liability mechanisms could help support a flourishing AI market, both by inspiring consumer confidence and ensuring better quality data. The European Commission's Expert Group on Liability and New Technologies53 advised that the liability rules remain broadly fit for purpose, but also set out several policy options:

• Identify operators. For many products, the operator is clearly identifiable. However, the users of AI products might in some cases have less control over its operation than other parties

such as service providers. This should be taken into account in identifying operators. • Make operators liable. Operators of high-risk AI tools could be subject to strict liability – that is, held responsible for damages resulting from their use even if no specific fault or criminal intent is identified, while operators of lower risk technologies remain responsible for the proper selection, operation, monitoring and maintenance of the technology. Whether risk is considered high or low could be determined by the severity and public reach of the risk. Responsibility for damage could be maintained regardless of how much autonomy is delegated to the AI. • Maintain responsibility for latent defects. Manufacturers could be held responsible for damage caused by defects even where these defects result from changes to the product that were within their control but took place after the item was put on the market. • Insurance

and compensation schemes. High-risk AI applications could be subject to mandatory insurance, much like private car ownership. Alongside this, compensation funds could be set up to compensate for damages that cannot be satisfied, for example because it was not possible to identify the party or technology responsible for damage.

##

• Assess data quality. Tools can be applied to identify quality issues such as incorrect data labels,
inappropriate biases, illegal material (including information gathered without consent) or 'fake
news'. Depending on the system, material might be removed automatically or flagged for
human review. The difficulty with such tools is that they are also biased, in this case against data
that does not conform to their definition of quality. Definitions of quality reflect perspectives
which are not always universally accepted. This is particularly true of such material as campaign
messages, news and misinformation. As such, these tools need to be developed cooperatively and continually tested to avoid manipulation or overuse. • Recognise the limitations. Data that reflects human decisions in domains that feature structural

biases cannot be complete, accurate and unbiased at the same time. In these cases, it is important to be aware of these biases and ensure that algorithms are not used in domains and functions for which they are not well suited.

4.2.4 Apply with care

There are technical reasons why AI should not be used to perform certain tasks. While AI can be good at pattern matching and identifying broad statistical correlations, it is not equipped to perform other tasks such as predicting individual social outcomes. Indeed, some of the most damaging examples of the misuse of algorithms come from the use of algorithms for tasks for which they are not well suited, such as predicting whether an individual will reoffend or perform well at work.62 On a wider scale, embedding AI-enabled systems in our infrastructures could introduce new vulnerabilities. At present, citizens are most directly exposed to functioning AI in content distribution, usually

designed to sell products and ideas. The case for promoting AI would be stronger if its development was mobilised to provide profound and tangible social good rather than minor efficiency gains, particularly when the costs and benefits are unevenly distributed. • Limit some technologies or application domains. Domains such as justice, policing and employment have been highlighted as inappropriate for the use of AI. However, not all AI applications within these domains are risky. Within justice, for example, there are many uncontroversial applications, such as supporting case law analysis or access to law. The European Commission for the Efficiency of Justice63 differentiates between uses that should be encouraged, that require considerable methodological precautions, which should be subject to study, and should only be considered with the most extreme reservations. Similarly, controversial AI techniques such as facial

recognition have been flagged as fundamentally
unacceptable in contexts such as mass identification in public places, but acceptable in others
such as identity verification to unlock phones.
• Adopt a risk-based approach. There are many ways of defining which applications are high risk
and what measures would apply in these cases.

##

operational, they need to be translated into specific measures, which may reveal how certain processes and applications diverge from principles and could lead to the collapse of consensus. To manage the shift from general to specific guidelines, ethicists could work with developers to explore the possibilities and examine their effects. As discussed in section 4.2.1., there are some good initiatives in this direction although their impact is limited by their voluntary nature. • Shift from voluntary to binding. Aside from those elements already established in law, adherence to ethics frameworks remains voluntary. When firms define their own codes, it is difficult to establish whether they make a substantial practical difference, and there are no mechanisms for enforcing their own compliance, or for ensuring that adopting

principles does not create a competitive disadvantage. If sufficient industry-wide principles cannot be achieved, binding legal measures can be developed. However, these would require even more specific interpretation of principles as well as penalties for noncompliance, raising the stakes and risking further degradation of consensus. It could help to reorient discussions about AI ethics to AI rights, as the latter already have legal force, albeit in general terms, and are often closely linked to ethical principles. • Establish digital ethics committees. Following the example of bioethics committees at institutional, national and international level, AI or digital ethics committees could be established to advise governments and provide an interface with international organisations, research councils, industry bodies and other institutions. • Integrate ethicists meaningfully. To ensure the conformity of products with

**ethical principles,
it has been suggested that AI ethicists could be embedded into firms and development teams. However, the effectiveness of such an approach depends upon the roles to which they are
assigned and the priorities of the activity. Ethicists can be employed for 'ethics washing' and
managing reputational risks, and their influence may be limited to 'low-hanging fruit' and 'win-
wins'. To succeed, ethicists would need to be deeply embedded in development teams and have
enough technical expertise and management support to make a difference. • Consider moratoriums carefully. Moratoriums have been suggested in response to ethical
issues presented by AI applications that are already in use (such as facial recognition), that are
technically feasible but not currently in use (such as fully autonomous lethal weapons), and that
are purely speculative (such as artificial consciousness). Temporary restrictions could**

allow time
to examine the impacts and options, while permanent bans aim to outlaw applications or stop
them from being developed in the first place. Whatever the approach, moratoria rely upon widespread consensus and trust.

##

information about an individual, such as a credit score, differs from bias that is based upon
statistical information about groups of people that are categorised as being similar. While embracing AI's discriminatory power, safeguards are needed to counteract the risk of reinforcing
and exacerbating undesirable social bias and inequality. These could include a combination of
technical, regulatory and social measures to better understand how algorithms make decisions,
the impacts of these decisions including their distributional effects, and mechanisms for reporting problems and challenging decisions.
• Avoid applications beyond AI's capabilities. Some applications of AI are predicated upon concepts that we know to be false. For example, facial categorisation technologies that claim to
be able to read emotions, identify sexuality, recognise mental health issues or predict

performance. The problem is not only that AI cannot perform these tasks accurately, but that the suggested relationship between the input and output lacks scientific credibility and, like eugenics, could provide a baseless veneer of objectivity for biased decisions and structural inequalities. Similarly, ML algorithms are more suited to finding trends and correlations than causal relationships. This means they can be useful for making predictions where relationships are straightforward, but less so at predictions about individual social outcomes within complex systems. The application and level of autonomy granted to AI should be guided by a sound understanding of what is scientifically credible and within the capabilities of today's AI. This is particularly important in key decision-making domains such as employment, insurance and justice. • Avoid applications with undesirable impacts. AI can be misused to predict

performance from facial images or individual social outcomes from statistical data. The use of such tools can lead to unequal distribution of impacts and deviate from established principles such as the presumption of innocence.86 Some AI applications that do perform well in their defined task can still be considered undesirable, such as personalised political advertisements in the context of election campaigns. The application and level of autonomy granted to AI should be guided by the understanding of factors beyond their direct aims, including their effectiveness and scientific basis, the wider impacts and their distribution, and their compatibility with social values. The development and use of algorithmic impact assessment could support this task. • Maintain human autonomy. Despite some debate over the details, there is a broad consensus

that humans should remain ultimately in control of AI. This may require some vigilance as the
detection of automation bias have shown human propensities to accept the advice of automated machines over that of humans, occasionally with tragic consequences, such as aircraft crashes.