acm Europe Technology Policy Committee

Association for Computing Machinery
acmpo@acm.org | @acmpolicy
www.acm.org/public-policy/europe-tpc

19 June 2020

Submitted Electronically

Mr. Roberto Viola, Director General
Directorate-General for Communications
Networks, Content and Technology
European Commission
Avenue de Beaulieu 25/Beaulieulaan 25
1160 Brussels, Belgium

Re: Comments on Commission's February 2020 AI White Paper (COM(2020) 65 final

Dear Director General Viola:

The Association for Computing Machinery (ACM) is the world's longest established and – with 17,000 members in Europe and 100,000 globally – the largest society of individual professionals engaged in computing in Europe and the world. ACM's Europe Technology Policy Committee (Europe TPC) is comprised of computer science experts from 10 countries. It is charged with providing European policy and law makers with expert and timely substantive input on computing technology, and the legal and social issues to which it gives rise. We are a non-profit and entirely nonpolitical organisation.

Consistent with its remit, Europe TPC is today pleased to submit its initial comments in questionnaire form on the Commission's February 2019 White Paper on *Artificial Intelligence - A European approach to excellence and trust* accompanied by, for the Commission's reference, by *When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*, the Committee's March 2018 White Paper co-authored with Informatics Europe. We also commend to your attention our *Statement on Algorithmic Transparency and Accountability*, produced jointly with ACM's U.S. Technology Committee in 2017.

ACM's Europe Technology Policy Committee stands ready to answer any questions about these comments and the technologies they address that you, your staff, or the Commission may have. To reach any of the Committee's or ACM's technical experts worldwide, kindly contact me directly at Oliver.Grau@gmx.net or ACM's Director of Global Policy and Public Affairs, Adam Eisgrau, at acmpo@acm.org.

Sincerely,

Oliver Grau, Chair
ACM Europe Technology Policy Committee

Europe TPC

Consultation Submission for EUC Whitepaper

Section 1 - An ecosystem of excellence

To build an ecosystem of excellence that can support the development and uptake of AI across the EU economy, the White Paper proposes a series of actions. **In your opinion, how important are the six actions proposed in section 4 of the White Paper on AI (1-5: 1 is not important at all, 5 is very important)?**

Item	Rating
Working with Member states	**5**
Focusing the efforts of the research and innovation community	**5**
Skills	**5**
Focus on SMEs	**5**
Partnership with the private sector	**5**
Promoting the adoption of AI by the public sector	**5**

Are there other actions that should be considered?

Yes. Working with general public is quite key. Also collaboration and partnership with organisations that represent professionals and other type of individuals would be quite key. These could encompass organisations like the ACM, but also standard bodies like the ISO, and open source foundations like the Linux Foundation and the NumFocus (which are the governing body of several key open source AI toolsets).

Revising the Coordinated Plan on AI (Action 1)

The Commission, taking into account the results of the public consultation on the White Paper, will propose to Member States a revision of the Coordinated Plan to be adopted by end 2020. **In your opinion, how important is it in each of these areas to align policies and strengthen coordination as described in section 4.A of the White Paper (1-5: 1 is not important at all, 5 is very important)?**

Item	Rating
Strengthen excellence in research	**5**
Establish world-reference testing facilities for AI	**5**
Promote the uptake of AI by business and the public sector	**4**
Increase the financing for start-ups innovating in AI	**4**
Develop skills for AI and adapt existing training programmes	**5**
Build up the European data space	**5**

Are there other areas that that should be considered?
[left blank]

<u>A united and strengthened research and innovation community striving for excellence</u>
Joining forces at all levels, from basic research to deployment, will be key to overcome fragmentation and create synergies between the existing networks of excellence. **In your opinion how important are the three actions proposed in sections 4.B, 4.C and 4.E of the White Paper on AI (1-5: 1 is not important at all, 5 is very important)?**

Item	Rating
Support the establishment of a lighthouse research centre that is world class and able to attract the best minds	**5**
Network of existing AI research excellence centres	**5**
Set up a public-private partnership for industrial research	**5**

Are there any other actions to strengthen the research and innovation community that should be given a priority?
[left blank]

<u>Focusing on Small and Medium Enterprises (SMEs)</u>
The Commission will work with Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI. **In your opinion, how important are each of these tasks of the specialised Digital Innovation Hubs mentioned in section 4.D of the White Paper in relation to SMEs (1-5: 1 is not important at all, 5 is very important)?**

Item	Rating
Help to raise SME's awareness about potential benefits of AI	**4**
Provide access to testing and reference facilities	**5**
Promote knowledge transfer and support the development of AI expertise for SMEs	**5**
Support partnerships between SMEs, larger enterprises and academia around AI projects	**5**
Provide information about equity financing for AI startups	**5**

Are there any other tasks that you consider important for specialised Digital Innovations Hubs?
It will be quite important to build the financial infrastructure to be able to back high-risk ventures given both the current economic climate together with the high risk nature of advanced research-based and highly innovative companies. If the ecosystem is not supportive towards SMEs and startups taking big leaps with high reward high risk opportunities the opportunity for innovations to come from Europe will be reduced.

Section 2 - An ecosystem of trust

Chapter 5 of the White Paper sets out options for a regulatory framework for AI. **In your opinion, how important are the following concerns about AI (1-5: 1 is not important at all, 5 is very important)?**

Item	Rating
AI may endanger safety	**5**
AI may breach fundamental rights (such as human dignity, privacy, data protection, freedom of expression, workers' rights etc.)	**5**
The use of AI may lead to discriminatory outcomes	**5**
AI may take actions for which the rationale cannot be explained	**5**
AI may make it more difficult for persons having suffered harm to obtain compensation	**5**
AI is not always accurate	**5**

Do you have any other concerns about AI that are not mentioned above? Please specify:
Cybersecurity concerns will be quite critical as new emerging technologies are introduced in production systems, new loopholes may also be introduced. The skills and best practices required to mitigate these risks will be critical to ensure high profile and high negative impact incidents don't take place due to security flaws of machine learning systems being exploited.

Do you think that the concerns expressed above can be addressed by applicable EU legislation? If not, do you think that there should be specific new rules for AI systems?
[*No opinion]*

If you think that new rules are necessary for AI system, do you agree that the introduction of new compulsory requirements should be limited to high-risk applications (where the possible harm caused by the AI system is particularly high)?
[*Other]*

If you wish, please indicate the AI application or use that is most concerning ("high-risk") from your perspective:
The compulsory requirements, and more importantly the level of scrutiny required should certainly be proportionate to the risk involved as suggested, however there should still be a compulsory step / process that must ensure this is evaluated upfront, and reasonable explanation can be provided (retrospectively) in the case it's decided the more involved "compulsory processes" are not required.

In your opinion, how important are the following mandatory requirements of a possible future regulatory framework for AI (as section 5.D of the White Paper) (1-5: 1 is not important at all, 5 is very important)?

Item	Rating
The quality of training data sets	**5**
The keeping of records and data	**3**
Information on the purpose and the nature of AI systems	**5**
Robustness and accuracy of AI systems	**5**
Human oversight	**5**
Clear liability and safety rules	**5**

In addition to the existing EU legislation, in particular the data protection framework, including the General Data Protection Regulation and the Law Enforcement Directive, or, where relevant, the new possibly mandatory requirements foreseen above (see question above), do you think that the use of remote biometric identification systems (e.g. face recognition) and other technologies which may be used in public spaces need to be subject to further EU-level guidelines or regulation:

[Other special requirements in addition to those mentioned in the question above should be imposed (please specify)]:

Facial Recognition is an area of AI technology where ADM (automated decision making) takes place about humans. It is therefore essential to understand the basic nature of ADM.

- *ADM is never neutral.*
- *The creators of ADM processes are responsible for their results. ADM processes are not only created by their creators.*
- *ADM processes must be traceable so that they can be subjected to democratic control.*
- *Democratic societies have a duty to create this accountability through a combination of technology, regulation and appropriate oversight institutions.*
- *We must decide how much of our freedom we want to transfer to ADM, and these discussions should be carried out through forums that provide the public, including minorities, a voice.*

The technology policy committees of the ACM are currently working on a position statement on Facial Recognition systems which we would be interested to share in more detail with the European Commission.

Do you believe that a voluntary labelling system (Section 5.G of the White Paper) would be useful for AI systems that are not considered high-risk in addition to existing legislation?
[*No opinion*]

Do you have any further suggestion on a voluntary labelling system?
Data labelling for AI data is often a task that may require specialised expert domain knowledge, or at the minimum structure control and processes (depending on what the exploratory or predictive analysis task is). Therefore a voluntary labelling system may not uphold the required standards required to ensure quality data, especially in use cases where data labelling would require a much larger upfront investment. Instead, having other ways where datasets could be scrutinised directly.

What is the best way to ensure that AI is trustworthy, secure and in respect of European values and rules?

[A combination of ex-ante compliance and ex-post enforcement mechanisms]

Do you have any further suggestion on the assessment of compliance?
Although the answer provided is "a combination of ex-ante compliance and ex-post enforcement mechanisms", this is something that would require industry- and even use-case-specific context requirements. In this situation industry standards, and consequently standardisation bodies, would play a key role, and should be engaged accordingly to ensure relevant and reasonable assessments processes and surveillance processes are established both internally and by external organisations.

Section 3 – Safety and liability implications of AI, IoT and robotics

The overall objective of the safety and liability legal frameworks is to ensure that all products and services, including those integrating emerging digital technologies, operate safely, reliably and consistently and that damage having occurred is remedied efficiently. **The current product safety legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. However, which particular risks stemming from the use of artificial intelligence do you think should be further spelled out to provide more legal certainty?**

[Mental health risks]

In your opinion, are there any further risks to be expanded on to provide more legal certainty?
The risks mentioned above are certainly critical, primarily due to the yet un-certain risks that AI poses in society and the individual. Particularly the mental health risks of the impact of these technologies is not yet fully understood, and this has led to loopholes in old regulatory frameworks to be exploited for corporate or personal benefit. It is critical to ensure mental health risks are one critical risk factor to be addressed and taken into consideration when designing, developing and operating AI systems.

Do you think that the safety legislative framework should consider new risk assessment procedures for products subject to important changes during their lifetime?
[*No opinion*]

Do you have any further considerations regarding risk assessment procedures?
[left blank]

Do you think that the current EU legislative framework for liability (Product Liability Directive) should be amended to better cover the risks engendered by certain AI applications?
[*No opinion*]

Do you have any further considerations regarding the question above?
[left blank]

Do you think that the current national liability rules should be adapted for the operation of AI to better ensure proper compensation for damage and a fair allocation of liability?
[*No opinion*]

Do you have any further considerations regarding the question above?
[left blank]

usacm ACM US Public Policy Council

acm Europe Council

Updated May 25, 2017

Statement on Algorithmic Transparency and Accountability

by ACM U.S. Public Policy Council, approved January 12, 2017
ACM Europe Policy Committee, approved May 25, 2017

Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment. The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset.

An algorithm is a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

There is also growing evidence that some algorithms and analytics can be opaque, making it impossible to determine when their outputs may be biased or erroneous.

Computational models can be distorted as a result of biases contained in their input data and/or their algorithms. Decisions made by predictive algorithms can be opaque because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations). Even well-engineered computer systems can result in unexplained outcomes or errors, either because they contain bugs or because the conditions of their use changes, invalidating assumptions on which the original analytics were based.

The use of algorithms for automated decision-making about individuals can result in harmful discrimination. Policymakers should hold institutions using analytics to the same standards as institutions where humans have traditionally made decisions and developers should plan and architect analytical systems to adhere to those standards when algorithms are used to make automated decisions or as input to decisions made by people.

This set of principles, consistent with the ACM Code of Ethics, is intended to support the benefits of algorithmic decision-making while addressing these concerns. These principles should be addressed during every phase of system development and deployment to the extent necessary to minimize potential harms while realizing the benefits of algorithmic decision-making.

Principles for Algorithmic Transparency and Accountability

1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

2. Access and redress: Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

3. Accountability: Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.

4. Explanation: Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.

5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

6. Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.

7. Validation and Testing: Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

When Computers Decide:

European Recommendations on Machine-Learned Automated Decision Making

Informatics Europe & EUACM
2018

INFORMATICS EUROPE

acm Europe Council

euacm ACM Europe Policy Committee

Report Authors

Professor James Larus
École Polytechnique
Fédérale de Lausanne

Chris Hankin
Imperial College London

Siri Granum Carson
Norwegian University
of Science and Technology

Markus Christen
University of Zurich

Silvia Crafa
Università di Padova

Oliver Grau
Intel

Claude Kirchner
Inria

Bran Knowles
Lancaster University

Andrew McGettrick
University of Strathclyde

Damian Andrew Tamburri
Jeronimus Academy of Data Science

Hannes Werthner
Technische Universität Wien

Sponsoring Organizations

Informatics Europe represents the academic and research community in Informatics in Europe. Bringing together university departments and research laboratories, it creates a strong common voice to safeguard and shape quality research and education in Informatics in Europe. With over 120 member institutions across 30 countries, Informatics Europe promotes common positions and acts on common priorities.

The **ACM Europe Council** aims to increase the level and visibility of Association for Computing Machinery (ACM) activities across Europe. The Council comprises European computer scientists committed to fostering the visibility and relevance of ACM in Europe, and is focused on a wide range of European ACM activities, including organizing and hosting high-quality ACM conferences, expanding ACM chapters, improving computer science education, and encouraging greater participation of Europeans in all dimensions of ACM.

The **ACM Europe Policy Committee (EUACM)**, a standing committee of the ACM Europe Council, promotes sound public policy and public understanding concerning a broad range of issues at the intersection of technology and policy. It does so by fostering dialogue and the robust exchange of ideas among the European Commission, governmental bodies in Europe, and the informatics and computing communities on the importance and impact of technology in many spheres. These include: job creation, economic growth, competition, investment, research and development, education and social development, and innovation.

Endorsements

This paper has been endorsed by the Informatics Europe Board, the ACM Europe Council, EUACM, and ERCIM.

Acknowledgements

Thanks to Fabrizio Gagliardi and Amalia Hafner for their inputs at all stages in the writing of this paper. Thanks also to Adam Eisgrau and his colleagues at ACM HQ for their editorial support.

Contents

Foreword

Over the past two decades, the ability of machines to challenge and beat humans at complex games has made "quantum" leaps, rhetorically if not in technical computing terms.

In 1997, IBM's Deep Blue supercomputer used "brute force" computing power to out-calculate Grand Master Garry Kasparov at chess. In 2011, the company's Watson employed "machine learning" (ML) techniques to beat several former *Jeopardy* champions at their own game. In early 2016, Google's DeepMind AlphaGo program—trained by a massive game history—repeatedly defeated the reigning European champion at *Go*: a game that has more possible board configurations than there are atoms in the universe [1]. It reached this milestone by employing two neural networks powered by sophisticated "automated decision making" (ADM) algorithms. And, in 2017, AlphaGo Zero became the strongest *Go* player on the planet—human or machine—after just a few months of game-play training alone. Incredibly, it was programmed initially only with the rules of the game [2].

Automated decision making concerns decision making by purely technological means without human involvement. Article 22(1) of the European General Data Protection Regulation (GDPR) enshrines the right of data subjects not to be subject to decisions, which have legal or other significant effects, being based solely on automatic individual decision making. As a consequence, in this paper we consider applications of ADM to applications other than those based on personal information, for example the game-playing discussed above. We discuss other aspects of GDPR later in the paper.

Whilst the game-playing results are impressive, the consequences of machine learning and automated decision making are themselves, however, no game. As of this writing, they have progressed to enable computers to rival humans' ability at even more challenging, ambiguous, and highly skilled tasks with profound "real world" applications, such as: recognizing images, understanding speech, and analysing X-rays among many others. As these techniques continue to improve rapidly, many new and established companies are utilizing them to build applications that reliably perform activities that previously were done (and doable) only by people. Today, such systems can both augment human decision making and, in some cases, replace it with a fully autonomous system.

In this report, we review the principal implications of the coming widespread adoption of ML-driven automated decision making with a particular emphasis on its technical, ethical, legal, economic, societal and educational ramifications. We also make a number of recommendations that policy makers might wish to consider.

Executive Summary

Computers process information and computers make decisions. Until recently, these decisions have been relatively simple. Are there sufficient funds in an account to permit a transfer? Which ads should be shown on a web page?

Substantial advances over the past decade in machine learning (ML) techniques, however, have produced systems that sometimes rival or even exceed human ability at challenging, ambiguous, and highly skilled tasks such as image and speech recognition, radiological image analysis, game play, and others. These techniques are still improving rapidly, but already many new and established companies are assembling them into applications that accomplish activities previously performed (and performable) only by people, such as driving vehicles, diagnosing illnesses, or even recommending judicial decisions. Collaborations between people and ML can enhance human acuity and actions, possibly resulting in fewer accidents and improved medical treatment.

Today, due to very recent breakthroughs in deep neural network (DNN) technology, these ML-powered automated decision-making (ADM) systems can both augment human decision making or, in some cases, replace it with a fully autonomous system, leading to potentially massive loss of jobs and de-skilling of the humans who hold them. Such systems are a focus of intense interest by the public, media and policy makers because ADM is the enabling technology for self-driving cars; robotic assistants; and the automation of many non-trivial human, government and commercial processes.

In contrast to traditional explicitly written computer programs, machine-learned systems are “trained” by exposing them to a large number of examples and rewarding them for drawing appropriate distinctions and making correct decisions. This distinction, while it may seem esoteric, has far-reaching consequences for our ability to understand the behaviour of these systems and for our confidence that they will behave consistently in an appropriate manner.

Forthcoming applications of ADM systems, most immediately and visibly self-driving cars, raise the possibility of large and potentially lethal physical objects operating under the control of ML models. Beyond the many ethical, legal, and practical concerns about turning over control and responsibility to machines, such systems also challenge the economic interests of people and enterprises who will compete with them. Concerns about ML also extend beyond physical systems, as other potential delegations of human judgment to machines—such as aiding judges in making sentencing and incarceration decisions in criminal actions—present serious issues about the equity and fairness of an opaque, inexplicable, and potentially biased process making life-changing decisions.

In addition, the capability of ADM systems raises serious ethical questions about whether these advances should be pursued at all. For example:

- Is it acceptable to permit machines to autonomously decide to kill humans, even if this decision is made in the context of a war?
- Do we believe that a machine should be empowered to make ethical judgments that have challenged philosophers for centuries, such as whether—faced with two horrible choices—a self-driving car should “choose” to hit an old pedestrian rather than crash into oncoming traffic, thereby risking the deaths of its more numerous and younger passengers?
- Or, to pick a less lurid example, can we train systems to produce results untainted by gender, race, class, and other bias when the training data used to “educate” these systems is produced by humans who share these biases to a greater or lesser degree?

As a practical matter, it is dangerous to “outsource” such ethical questions related to ADM systems to expert committees or to industry. Rather, they require a deep understanding and incorporation of ethics throughout the design of the technology. Social and moral values thus should no longer be seen simply as "risk factors" or constraints, but as primary drivers and shapers of innovation.

Recommendations

Technical

1. ***Establish means, measures and standards to assure that ADM systems are fair.*** All key actors—academia, industry, government institutions, international institutions, NGOs, and citizens—must be involved in the formulation of standards and practices that ensure that the public good is the primary criterion for assessing ADM quality. These standards need to be broad and principled to stay relevant to the rapidly evolving technology and industrial applications of ADM. To facilitate this objective, research is needed to develop a solid theoretical basis for machine learning and techniques for explainable automated decision making.

Ethical

2. ***Ensure that Ethics remain at the forefront of, and integral to, ADM development and deployment.*** As with health and biology, member countries as well as the European Union should develop ethics committees to advise the societal, political, academic, and legal systems about the positive as well as negative consequences of ADM initiatives, tools, and systems. As a guardian of the public interest, a new European agency could oversee the development and deployment of machine-learned ADMs throughout Europe.
3. ***Promote value-sensitive ADM design.*** Appropriate programmes throughout higher education should teach techniques such as value-sensitive design and otherwise stress that social values and the ethical priorities of technology users must be designed into all aspects and elements of ADM.

Legal

4. ***Define clear legal responsibilities for ADM's use and impacts.*** The core principles currently governing ADM development within the computing professions—accountability, traceability, and responsibility—should be adopted as the basis for broad discussion and debate among legal and technical experts, the media and society at large in pursuit of new legal norms to govern wide-scale ADM deployment. In particular, the blanket disclaimer of liabilities attached to virtually all software today should be revisited and revised or rejected if, as it appears, it is inapplicable to many current and likely uses of ADM. The EU agency proposed in Recommendation 2 should foster and facilitate this debate and recommend responsive legislation as and when appropriate.

Economic

5. ***Ensure that the economic consequences of ADM adoption are fully considered.*** Among its first official acts, and for the purpose ultimately of issuing appropriate guidelines and regulations, the new Agency proposed above might productively solicit immediate comment on a range of defined economic and socio-economic issues to which the accelerated development and application of ADM likely will give rise. Its permanent agenda should explicitly be acknowledged

to consist of two, inherently interrelated goals: fostering the responsible evolution and use of ADM systems and minimizing the resulting personal, societal and economic disruptions to individuals and nations.

Societal

6. ***Mandate that all privacy and data acquisition practices of ADM deployers be clearly disclosed to all users of such systems.*** Data is the fuel for machine learning. Where and whenever information is collected, what is being collected and the uses to which it will be put should be described to the data provider concisely and clearly.
7. ***Increase public funding for non-commercial ADM-related research significantly.*** Additional research is necessary to better understand machine learning and its use in systems to influence human behaviour. Many fundamental issues remain to be investigated. Robust public knowledge of these techniques, without depending predominantly upon industry for research results, is a prerequisite for a broader debate about their acceptability as well as for effective and principled adoption of these techniques by European companies. Improved techniques for explainable automated decision making should be a research priority.

Educational

8. ***Foster ADM-related technical education at the University level.*** All university students should receive instruction in the practicalities and potential of machine learning. Students of all disciplines need to be aware of the impact this technology will have on their field and future work.
9. ***Complement technical education with comparable social education.*** Because of the increasing impact that technology will have on society, technical curricula also should educate students to deal with complex scenarios by complementing technical skills with the development of critical thinking, digital wisdom, and ethical judgement. Higher education curricula should foster inter-disciplinary studies, drawing from the European cultural heritage in both scientific disciplines and liberal arts. An accessible introduction to ADM and the issues that it raises also should be introduced into secondary education curricula.
10. ***Expand the public's awareness and understanding of ADM and its impacts.*** There is a clear need to educate the general public in this technology, as it is being rapidly introduced and will affect virtually everyone in their professional and private lives. Since most people do not take additional courses after completing their formal education, the public media thus represent the broadest de facto means of educating the general population. Accordingly, information and discussions of the type contained in this paper must be presented to the press by computing professionals and technology policy makers. Due consideration must be given to the troubling use of ML techniques to shape public opinion.

Introduction

Technological innovation is rarely smooth and continuous. Years of slow, incremental research and development precede the seemingly instantaneous introduction, adoption, and deployment of a new technology, which in turn disrupts long-standing interpersonal, societal, political, and economic relationships. We have reached such a point with machine learning (ML) systems.

This white paper is focused on machine learning because ML is the fundamental technology underlying a broad array of emerging products and services that are popularly grouped under the rubric of Artificial Intelligence (AI). The social and economic concerns raised for decades about Artificial Intelligence (AI) thus are now questions and concerns about ML, including whether artificially intelligent machines and robots will rapidly surpass, supplant and displace humans socially and especially economically [3] [4] [5].

From a technical perspective, machine learning systems, in contrast to explicitly written programs, are "trained," by exposing them to a large number of examples and rewarding them for drawing appropriate distinctions and making correct decisions, much in the same way as human beings learn. This distinction, while it may seem esoteric, has far-reaching consequences for our ability to understand the behaviour of these systems and for our confidence that they will behave in an appropriate manner.

From an economic perspective, a key distinction of systems controlled by ML is that they provide solutions to problems that are very difficult to express, let alone solve with a conventional computer program. In writing a program, a software developer must anticipate the multitude of possible scenarios that might be encountered by the program and explicitly develop a response for each. The abundance of bugs in programs illustrate the difficulty of this process, even in simple domains. More complex problems, even if they are skills that young children acquire easily, such as vision, speech, and navigation, have proven difficult if not impossible to analyse and express in this manner.

Machine learning has been studied and applied for many years. However, within just the past 5 years, breakthroughs in the application of deep neural network techniques have radically increased the speed and accuracy of automated decision-making systems. This breakthrough was made possible by improved algorithms for training deep neural networks (DNNs), as well as dramatic increases in computation now viable with GPU coprocessors and cloud computing. It was further facilitated by advances in the ability to collect and store the huge amounts of data typically needed to train these systems.

Such systems are a focus of intense interest because machine learning (ML) increasingly is the enabling technology for automated decision making (ADM), which lies at the heart of technologies that, especially in recent years, have galvanized the attention of the media, public and policy makers. These notably include self-driving cars, robotic assistants, package-delivering drones, facial and speech recognition, and an ever-expanding array of other non-trivial human, commercial and governmental processes to which DNN has brought human-level performance.

Today, fuelled by important and very recent technical breakthroughs involving deep neural networks, researchers, industry and even governments are racing to apply the latest machine learning systems in myriad domains. Leading technology companies—such as Alibaba, Amazon, Baidu, Facebook, Google, Microsoft, and Tencent—are rebuilding their internal infrastructure and product lines around machine learning. Indeed, China has identified further progress in machine learning as the linchpin of its goal to globally dominate technology by 2025 [6].

These activities, in turn, raise broader questions and concerns as ML-based systems rapidly enable greater and greater degrees of automated decision making and are being deployed to augment or substitute for human intelligence, judgement and effort. For example, while self-driving cars are both potentially convenient and productivity-boosting, they also may be viewed as potentially lethal physical objects autonomously operating under the control of ML-enabled systems.

Beyond such legal and practical concerns, turning over control and responsibility to machines also raises many important macro- and micro-economic concerns for both society and individuals, including people who make their living driving cars, taxis, and trucks and the companies that currently employ them. Finally, as ML's applications extend beyond physical systems into more metaphysical pursuits—such as "advising" judges in making sentencing and incarceration decisions in criminal actions—serious ethical concerns arise about the equity and fairness of an opaque, inexplicable, and potentially biased process helping to make such radically life-changing decisions. [7]

The pace of developments in this sphere in recent years—and the certain acceleration of that pace in the future due to natural maturation of this leading-edge technology, competitive pressures, and the education of many skilled practitioners—makes an understanding of the strengths and weaknesses of ML essential to understanding the impact of this trend and its consequences.

This white paper, a joint product of Informatics Europe and EUACM (the ACM Europe Policy Committee), is intended to contribute to that understanding by broadly highlighting what is possible and likely in the future. It also presents specific recommendations from the European technical and scientific community about how policy makers, legislators, and concerned individuals might best respond to the rapid growth of ADM. These are grounded in a presentation of key technological background and concerns about these remarkable and revolutionary systems.

Machine Learning

Machine learning is not a new technique, but with the advent of increasingly powerful computers and vast amounts of data, it has been applied to far more complex and less well-structured problems than were traditionally targeted by statistics, ML's intellectual predecessor.

Machine learning constructs models from data. These models in turn can be used to make predictions about values that were not part of the original training set. Consider the simplest form of machine learning, linear regression, which predicts the value of a (dependent) variable based on the value of a second (independent) variable assuming a linear relationship between them. For example, the independent variable can be mileage of a used automobile and the predicted value its sale price. The model can be visualized as the line in a chart that minimized the distance from the known data points, as illustrated in Figure 1. This line can be used to make predictions about the sale of other cars.

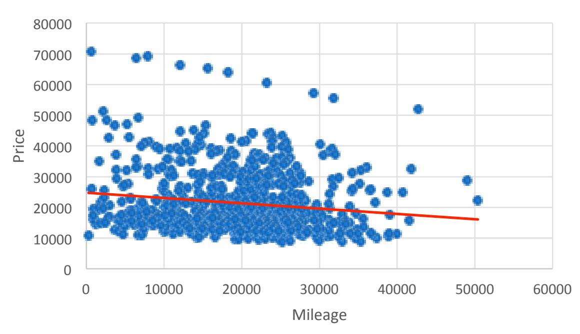


Figure 1 Used car price and mileage (from [8]).

This basic idea can be expanded in several directions. Instead of using a straight line, we could choose a higher dimensional curve that better fits the data. Continuing with our example, at some point, the sale price of a car might decrease faster as its mileage increases. And, instead of just using one input variable (e.g. mileage), we can use a number of them (e.g. make, model, type, engine size, etc.) to make increasingly accurate predictions. It becomes difficult to depict the resulting functions in a simple picture, but techniques developed by mathematicians and statisticians have long been used to build these simple statistical models, which are ubiquitous.

Machine learning can be implemented in a number of different ways. In **supervised** machine learning, the algorithm is "trained" against a set of labelled data: for example, images with a human-written description of their content. In contrast, **unsupervised** machine learning aims to automatically extract features from the data using clustering algorithms: for example, recognizing that a group of images contain cats. An intermediate approach taken in some agent-based systems is **reinforcement** learning, which was used to train a champion ***Go*** playing program by having it play millions of games, observe the outcomes of its moves, and seek better future results for the functions analysed.

In **off-line** systems, the training occurs before the algorithm is applied to test data in a live context. In **on-line** systems, some prior training is conducted but the algorithms are continuously improved based on the live data.

So-called **deep learning** algorithms are examples of learning systems that use multiple levels of representation. Deep learning algorithms use multi-level neural networks where one or more of the levels are hidden. A recent Nature paper gives detailed information [9].

Machine-learned models are capable of producing accurate and subtle results if they are trained with large amounts of appropriate data. This flood of training data is housed in vast data centres and made possible by abundant and inexpensive computation and storage. Much of the data originates in online services, which monitor and record vast amounts of information about their users. Development of these processing techniques and infrastructure was motivated by the very profitable business of online advertising. They are far more generally applicable, however, and can be used to develop models of the physical world and individuals, so long as adequate data is available.

Implications of ADM Systems

By themselves, machine learned models are interesting intellectual enterprises. After being trained, they can make predictions, often with a surprising degree of accuracy. Of greater importance, however, is that these predictions provide a basis for making decisions that drive the behaviour of complex systems that interact with humans. For example, a machine-learned image recognizer can discern a stop sign in a cluttered landscape and a machine-learned control system can bring a car to a stop at the appropriate position.

On the positive side, advances in machine learning mean that valuable and useful tasks can now be performed for the first time by machine. These include real-time language translation, facial recognition for security, and autonomously driving a car (and other forms of transport). Medical advances include the recognition of the onset of diseases such as Alzheimer's several years before symptoms appear [10], or improving the accuracy of cancer diagnoses. Collaborative systems based on ADM can enhance human acuity and actions, possibly resulting in fewer accidents and improved medical treatment.

However, at the same time, applications of ADM will have far-reaching implications on society. These range from new questions about the legal responsibility for mistakes committed by these systems to retraining for workers displaced by these technologies.

Automated driving has become a lively focal point in the discussion about machine learning, automated decision making, and AI in general. Automobiles are not only complex and potentially deadly machines that we all encounter daily, but cars are also a means of individual mobility that address an important human need. In Western Europe, the average number of private cars is about 0.6 per person, so it is not surprising that there is a high level of interest in autonomous vehicles.

Automated driving is certainly a major technological milestone, but in fact its building blocks have been under development for some time and, to a limited extent, are available in other products. In vehicles, six levels of automation are distinguished.

They range from level 0, completely human driving, to level 5, completely driverless. Today many assistance functions (level 1) and some partial automation systems (level 2) are available on the market. Products at level 3 have been introduced. They can drive partially autonomously, but the driver is expected to maintain responsibility and be able to take over the control at any time. Under development are level 4 and 5 systems that are able to completely control the vehicle. The difference between these is that at level 5 the vehicle could be completely unmanned, for example, automated transport of goods and passengers. This is often referred to as autonomous driving. In these cases, a machine learned system will make decisions to control the vehicle based on sensing the local environment and from maps of it.

If AI and ML control a car, the established system of liability in case of accidents and fatalities will need to evolve. Today, by legislation, the final liability lies with a car's driver. As accidents are usually complex, drivers involved in an accident are held responsible if it can be proved that they made an avoidable mistake. Car manufacturers are responsible for technical functionality, e.g., they are liable if a braking system had a fault caused by a violation of due diligence obligations during design or manufacture. In a level 4 or 5 self-driving car, who or what will or should be responsible for an accident? The car's manufacturer? The creator of the control system? How can we determine if either or both parties fulfilled their due diligence obligations in constructing this system?

Beyond this set of concerns, the advent of self-driving cars will have significant economic consequences. Entire careers, such as taxi or delivery driver, will largely disappear. Many predict that privately owned cars will become rare as driverless services satisfy most peoples' need to use a car between 1-2 hours per day [11]. Furthermore, since the automotive industry employs a large labour force, 5.7% of employees in Europe, the consequences for companies, employees and even nations may be large [12]

Considerations for ADM Systems

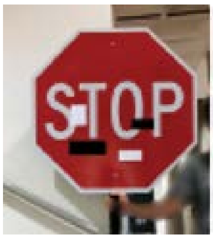
Technical

Machine learning systems in many (but not all) cases are "black box" systems in which data is presented to a model. It then produces predicted outcomes or recommendations without providing a tangible or verifiable explanation of how the outcome was reached. While conventional computer applications may appear to behave similarly, they have an internal logic and are constructed out of abstractions that make an application's logic and behaviour comprehensible and reliably predictable to its software developers. Whilst many individual algorithms have clear mathematical foundations, an underpinning theory or a science of machine learning does not yet exist, but one would be beneficial in understanding how to better train these systems in predicting their behaviour.

For many ML models, in particular deep neural nets, inexplicability is fundamental. Currently there is no underlying theory that explains how or why the models are effective for a particular type of problem and no baseline from which to predict their eventual performance. An ML model developer starts with a vast amount of data and performs a large computation to adjust the model parameters to produce the best predictions for this data, and then repeats this optimization process on different data collections until a satisfactory prediction accuracy is achieved.

The resulting models are equations, which compute mathematical functions with millions of input parameters [13]. Unlike equations used in the sciences, where the mathematics is used to describe the physical world, ML models are equations that have no obvious underlying physical or logical basis. Reading these models thus provides little insight into the underlying phenomena, where they originated, or how they will behave in a given situation. It is necessary to run a model to determine its effect. And, like other digital computations, there is no guarantee of consistency: a model may produce radically different results for two scenarios that seem quite similar to humans.

For example, this picture of a stop sign, slightly defaced as shown, was interpreted by a ML system as a 45-mph speed limit sign [14].



Humans have no problem recognizing the defaced sign, but ML systems are still far less able than we are to robustly tolerate ambiguity.

As a result, machine learning models are currently difficult to test adequately and rather easy to deceive or confuse, which presents challenging new security problems. Conventional software testing techniques rely heavily on unit tests, which validate small, individual components of a software system before they are combined into a unified system, which can then be tested as a complete unit. ML models, by contrast, are black boxes that can only be tested as a unified whole. Because they typically have a large number of inputs, it is not possible to thoroughly test even simple models, which leaves open the question of how a ML model will perform in a given situation. More subtle techniques modify images in imperceivable ways that confuse current ML classification systems [14].

Another serious concern is algorithmic bias, which arises when the training data or training process for a ML system introduces unwanted or illegal biases. The ACM "Statement on Algorithmic Transparency and Accountability" proposes 7 principles to alleviate this problem. [15]

While it is not clear that the General Data Protection Regulation (GDPR) imposes such a requirement on ADM systems or their creators [16], there has been research on constructing models that are explicable, so that the basis for their decisions can be presented and explained. For example, one technique known as "unconditional counterfactual explanation" describes the smallest change in the world that can be made to achieve a desirable outcome [17]. A simple example of a counterfactual explanation would be "You were denied a loan because your annual income was £30,000. If your income had been £45,000 you would have been offered a loan." The method allows us to identify changes to variables that would result in a different decision, which can be shared with the affected party.

RECOMMENDATION 1: ***Establish means, measures and standards to assure that ADM systems are fair.*** All key actors—academia, industry, government institutions, international institutions, NGOs, and citizens—must be involved in the formulation of standards and practices that ensure that the public good is the primary criterion for assessing ADM quality. These standards need to be broad and principled to stay relevant to the rapidly evolving technology and industrial applications of ADM. To facilitate this objective, research is needed to develop a solid theoretical basis for machine learning and techniques for explainable automated decision making.

Safety of Automated Driving

Acceptance of automated driving is a complex matter, involving many considerations. A majority of the public, the users of the technology, will accept automated driving only when they are convinced that it is safe. Yet there is no metric nor established evaluation procedure to demonstrate the fitness of the technique. Moreover, there is a widespread belief that technology is, or should be, perfect, even when it replaces demonstrably imperfect humans. In particular, if the goal is to reduce the number fatalities or accidents, it is difficult to demonstrate this before the introduction and widespread deployment of the technology. This presents one of the biggest hurdles for manufacturers as well as authorities. One viable alternative is the introduction of a standard catalog of test scenarios for evaluating automated driving systems. This catalog does not yet exist, and moreover it is not yet clear who would set it up and maintain it or how to demonstrate its value to the broad public.

Also, members of the public will need to consider their relationship to the new technology. When is it acceptable, *or reliable enough*, to use? What happens to my personal preferences? Will the machine react in my interest? Some scientific experiments studied personal choices and preferences (altruistic vs. egoistic). Initial results show that many individuals would not like to use a car that could kill them to save others [26]. These experiments argue that the public needs to be involved in defining the "correct" response to challenging situations.

Even short of a fully autonomous car, the connected nature of an automated car raises privacy issues. An automated and connected car is not just recording exact positional data, but it also captures driver interventions and all decisions made by the driver. The obvious questions are who would be in control of this information (driver, car manufacturer, authority), for what additional purposes will this data be used, and can the driver object to the collection and use of this data for objectionable purposes?

Ethical

The advent of ADM systems raises serious ethical questions about whether these technical advances should be pursued at all. To pick an extreme example, is society ready for machines that autonomously decide to kill a person, even if this decision is made in the context of a war? Or, do we believe that a machine should be empowered to make ethical judgments that have challenged philosophers for centuries, such as whether it is better that a self-driving car hit an old pedestrian rather than crash into oncoming traffic, thereby killing the more numerous and younger passengers in the car? Or to pick a less lurid example, can we train systems to produce results untainted by gender, race, class, and other bias when the data used to train these systems is produced by humans who share these biases to a greater or lesser degree [18]?

On a technical level, ambiguous questions pose great challenges for machine learning. Difficult ethical situations are often relatively uncommon and so will be infrequent in training data. The developers of a ML model must also be careful and thoughtful in labelling the data, identifying and resolving ethical dilemmas such as those above. And, despite a high level of care, there will still be a serious level of concern as to whether the ADM system will behave "ethically" in a given situation. Testing alone will not suffice. As the defaced stop sign example above demonstrates, minor changes in the perceived world could unpredictably change an appropriate response into an entirely inappropriate one, with possibly lethal consequences.

As a practical matter, it is dangerous to "outsource" the ethical questions related to ADM systems to expert committees, although they have an important role to play. These issues require a deep understanding of ethics throughout the design of the sociotechnical systems. Therefore, as discussed below, technical education and public education must emphasise the importance of such societal issues. Social and moral values must no longer be seen simply as "risk factors" or constraints, but as drivers and shapers of innovation. (For example, the idea that applications are developed in order to serve and support values and serve the interest of society has been further advanced under the rubric of Value Sensitive Design (VSD) in the work of Batya Friedman and others [19]. VSD explores how direct and indirect stakeholders are affected by technology in development through a combination of conceptual, empirical and technical investigations. More generally, the work conducted under the general heading of Human-Centered Computing, including interaction design, human-computer interaction and social computing, provides methods and techniques to combine human values and skills with computer capabilities.)

RECOMMENDATION 2: ***Ensure that Ethics remain at the forefront of, and integral to, ADM development and deployment.*** In a similar manner to health and biology, member countries as well as the European Union should develop ethics committees to advise

The German BMVI Ethics Commission

A European activity to mention in this context is the 'Ethics Commission on Automated and Connected Driving', which was appointed by the German Federal Minister of Transport and Digital Infrastructure (BMVI). The commission started its work in September 2016 and presented their report June 2017. [21] The report sets 20 ethical rules for the adaptation of automated and connected driving. Furthermore, it contains a discussion of open and unresolved issues.

The Commission was led by a former German Federal constitutional judge and included experts from philosophy, jurisprudence, social sciences, technology impact assessment, the automotive industry and software development. Their report does not contain radical new considerations, but it is an initiative by a public body to create a framework of ethical and related questions relevant for the introduction of automated driving. The report summarizes that automated driving can only be justified if it (significantly) reduces the harm to humans caused by traffic accidents. It also emphasizes broad societal values, like the protection and development of the individual in society. It also raises issues like data protection in the case of connected systems able to record the location and actions of drivers.

The report does not give hard guidance, but clearly states that automated driving should be used for the betterment of humankind and not be discriminatory in any way. The report also lists relevant technical requirements concerning matters like testability, security against attacks, privacy and how to switch between automated and driver control.

the societal, political, academic, and legal systems about the positive as well as negative consequences of the digital automation initiatives, tools, and systems. As a guardian of the public interest, a new European agency could oversee the development and deployment of machine-learned ADMs throughout Europe.

RECOMMENDATION 3: *Promote value-sensitive ADM design.* Appropriate programmes throughout higher education should teach techniques such as value-sensitive design and otherwise stress that social values and the ethical priorities of technology users must be designed into all aspects and elements of ADM.

Legal

In the end, powerful techniques like ADM that affect human life and economic welfare will be regulated and adjudicated by the legal system. Because legal and regulatory systems differ, it is not appropriate to be too specific in this document. These political and societal decisions, however, should be informed by a solid understanding of the technology and its application to the physical world. Some of the legal issues that legislators and courts prudently might address are:

- Who is liable for damage caused by autonomous systems? Should the same standards apply to open source software?
- Do such systems always act on behalf of third parties, like a proxy, or do they act on their own behalf?
- Who owns goods produced by (or with the authorization of) ADM, such as inventions or works of intellectual property?
- Can autonomous systems be understood as independent actors? Can they enter into legally valid contracts and obtain partial legal capacity or even full legal personality? Should ADM systems be taxed?
- Do we need standards to ensure that ADM systems are always recognizable as such by people?
- Can autonomous systems make themselves liable to prosecution if they take actions that are relevant under criminal law? How should such penalties be structured?
- Can autonomous systems rely on the protection of fundamental rights, *e.g.* freedom of expression and information, or economic freedom?
- How should the use of autonomous weapons systems be regulated under international law?

The following precepts drawn from current software governance models and practice may be useful in beginning to formulate answers to these and other complicated questions raised by the anticipated ubiquity of ADM systems. When complex automated decision-making systems operate without human involvement, three conditions need to be satisfied:

- **Accountability**—the creators, manufacturers, operators, maintainers, and users of automated decision-making algorithms and systems each must be accountable for relevant aspects of the process.
- **Traceability**—the automated decision maker must be able to autonomously show the rationale for its decision using empirical, experimental or other evidence that rejects several other possible decisions in favour of the one chosen.
- **Responsibility**—the automated decision maker and its creators, manufacturers, operators, maintainers, and users must be socially and legally responsible for its decisions.

These criteria are also appropriate to assessing “shared decision making,” in which humans and machines are both involved in producing an action and outcome.

Finally, in this context, it appears clear that ADM applications will require re-examination, and potentially revision or rejection, of the blanket disclaimer of liabilities attached to virtually all software today. It is one thing for a videogame producer to disclaim all liabilities for its software, and quite a different thing for the maker of a self-driving car to attempt to do the same for its auto-piloting system.

The eventual evolution of deployed automotive ADMs from level 3 to levels 4 and 5 (fully automated vehicles) will bring about a major change in the liability and raises a number of ethical issues that need to be addressed and resolved. Now software and hardware using AI and machine learning techniques are making decisions that can potentially cause harm to humans. New arrangements between the automotive industry, car owners and insurance companies need to be made. This also requires that new policies and legal frameworks be agreed to on an international level. Car makers and technology providers should be required to prove rigorously that their technology is safe under the specified conditions of use. [20, 21].

RECOMMENDATION 4: ***Define clear legal responsibilities for ADM's use and impacts.*** The core principles currently governing ADM development within the computing professions—accountability, traceability, and responsibility—should be adopted as the basis for broad discussion and debate among legal and technical experts, the media and society at large in pursuit of new legal norms to govern wide-scale ADM deployment. In particular, the blanket disclaimer of liabilities attached to virtually all software today should be revisited and revised or rejected if, as it appears, it is inapplicable to many current and likely uses of ADM. The EU agency proposed in Recommendation 2 should foster and facilitate this debate and recommend responsive legislation as and when appropriate.

Economic

ADM systems have the clear potential to perform a broad range of tasks with higher accuracy and at lower cost than now accomplished, or achievable, by humans. Such systems thus can benefit individual companies, national economies and society in general. Foreseeably, however, they also will place disproportionate and significant burdens on individuals displaced by ADM and those who depend upon them. Retraining and other support of such workers and households, in turn, likely will result in significant costs to the societies and economies of impacted nations. Comprehensive and coordinated discussion of such "external" human and economic costs, and the most effective policies and mechanisms for mitigating them, must be undertaken throughout Europe and, indeed, globally.

From a business and financial perspective, who will benefit from the advances in ADM technology and its practical applications? It is easy to point to numerous examples of widespread societal benefits from recent technology innovation, such as near-zero cost communications throughout the world or access to unprecedented amounts of knowledge and entertainment. Technological innovation over the past few decades also has produced significant wealth for a group of highly trained people and enormous wealth for a very small group of people.

Its effect on wealth distribution as a whole, however, is less clear. For example, in 2016, the US industrial production was at a record level, 47% higher than 20 years earlier, but it was accomplished with 29% less labour because of automation [22]. Computers and the Internet have nearly eliminated or redefined entire categories of jobs such as secretary and travel agent. Automated decision making, driven by ML, will allow even more jobs to be automated, including increasingly highly skilled, currently high paying positions.

Shifts of this type have occurred before. Historically the automation of farm labour by mechanization caused job losses on a huge scale, but other jobs emerged. Relatively few people would trade their current life and profession for the uncertainty and manual labour of their ancestors' farms. Nevertheless, the transition from farm to city, from farming to industry, was immensely painful for those involved in it and produced political turmoil in the 19th and 20th century in numerous countries. History eventually will show whether the adoption of computers, the rise of the Internet, and ADM constitute an equally significant social revolution. Even if they do not, the impact on millions of workers of the transformations that these technologies set in motion will continue to be profound.

RECOMMENDATION 5: ***Ensure that the economic consequences of ADM adoption are fully considered.*** Among its first official acts, and for the ultimate purpose of issuing appropriate guidelines and regulations, the new Agency proposed above might productively solicit immediate comment on a range of defined economic and socio-economic issues to which the accelerated development and application of ADM likely will give rise. Its permanent agenda should explicitly be acknowledged to comprise two, inherently interrelated goals: fostering the responsible evolution and use of ADM systems and minimizing the resulting personal, societal and economic disruptions to individuals and nations.

Societal

Even in a world in which ADM systems consistently function almost flawlessly, inevitably their deployment will have significant and often negative unintended consequences. Perhaps the clearest and best investigated examples of such problems come from experience with flight control systems in commercial airplanes. Such systems fly and land planes with great reliability and accuracy, often leaving pilots with little to do but monitor the plane's activity. Investigations of infrequent aviation accidents have found that, in many cases, pilots are unable to intervene effectively when something unexpected occurs because they do not have enough contextual knowledge to make a quick and appropriate decision. This has led to fatal accidents [23]. Indeed, Google's experience with self-driving cars has led the company to propose removing all controls for occupants of the car, who often are similarly distracted and/or unable to respond quickly to unexpected events [24].

Today, Facebook's ranking algorithm for an individual's "news feed" raises similar concerns about the effect on humans of taking us "out of the loop." Until recently, news was broadly distributed, and it was challenging to target narrow groups of like-minded individuals. Consequently, almost everyone was exposed to a variety of opinions, forcing individuals to actively select or reject information sources and points of view. With algorithm processes, human judgment is no long necessary or desirable.

Facebook's machine learned algorithm for selecting news feed items creates an "echo chamber," in which an individual's expressed interest in a topic or political point of view enables Facebook to group him or her with other like-minded individuals. This increases the person's engagement (time spent on Facebook), but filters out competing views [25]. In a filter bubble of this sort, it is easy to believe that everyone shares your views, with a consequential decline in critical thinking.

Moreover, numerous successful companies have demonstrated that people are willing to trade access to personal information (privacy) in return for "free," advertising-supported entertainment or services. This exchange is poorly understood by the participants. That may be a relatively benign result in a commercial context. Recent history demonstrates, however, that this type of advertising also can be effective at influencing political events.

What are the appropriate limits to using machine learning to influence human behaviour and societal direction? At this point, there are few if any legal or societal constraints on the use of these powerful techniques.

RECOMMENDATION 6: ***Mandate that all privacy and data acquisition practices of ADM deployers be clearly disclosed to all users of such systems.*** Data is the fuel for machine learning. Where and whenever information is collected, what is being collected and the uses to which it will be put should be described to the data provider in simple terms.

RECOMMENDATION 7: ***Increase public funding for non-commercial ADM-related research significantly.*** Additional interdisciplinary research is necessary to better understand machine learning and its use in systems to influence human behaviour. Many fundamental issues remain to be investigated. Robust public knowledge of these techniques, without depending predominantly upon industry for research results, is a prerequisite for a broader debate about their acceptability as well as for effective and principled adoption of these techniques by European companies. Improved techniques for explainable automated decision making should be a research priority.

Educational

In the near future, perhaps sooner than we think, virtually everyone will need a basic understanding of the technologies that underpin machine learning and artificial intelligence. This knowledge will enable people to most productively engage with intelligent devices and services that they encounter, purchase, and use.

Most fundamentally, people should understand that ADM systems differ fundamentally from prior computer applications. Automated decision-making systems will make mistakes. The assumption that computers are accurate and nearly infallible, while generally appropriate for tasks such as bookkeeping, is dangerously incorrect for ADM systems. These systems will outperform humans sometimes and will fail spectacularly at others. Effectively designed education campaigns and curricula can and should provide all individuals with a nuanced understanding of what these systems are capable of and how they might go wrong.

Broad diffusion of ADM technology is changing research in many disciplines, ranging from the sciences to the social sciences and humanities. Given the potential impact of machine learning on a wide range of disciplines, and its role in innovation and discovery, universities should teach courses in machine learning.

On the other hand, because of the increasing impact that technology will have on society, technical curricula should also educate students to deal with complex scenarios by complementing technical skills with the development of critical thinking, digital wisdom, and ethical judgement. Higher education curricula should foster interdisciplinary studies, drawing from the European cultural heritage in both scientific disciplines and liberal arts.

RECOMMENDATION 8: ***Foster ADM-related technical education at the University level.*** All university students should receive instruction in the practicalities and potential of machine learning. Students of all disciplines need to be aware of the impact this technology will have on their field and future work.

RECOMMENDATION 9: ***Complement technical education with comparable social education.*** Because of the increasing impact that technology will have on society, technical curricula also should educate students to deal with complex scenarios by complementing technical skills with the development of critical thinking, digital wisdom, and ethical judgement. Higher education curricula should foster interdisciplinary studies, drawing from the European cultural heritage in both scientific disciplines and liberal arts. An accessible introduction to ADM and the issues that it raises also should be introduced into secondary education curricula.

RECOMMENDATION 10: ***Expand the public's awareness and understanding of ADM and its impacts.*** There is a clear need to educate the general public in this technology, as it is being rapidly introduced and will affect virtually everyone in their professional and private lives. Since most people do not take additional courses after completing their formal education, the public media thus represents the broadest de facto means of educating the general population. Accordingly, information and discussions of the type contained in this paper must be presented to the press by computing professionals and technology policy makers. Due consideration must be given to the troubling use of ML techniques to shape public opinion.

Conclusion

Exponential growth in the sophistication and ubiquity of ADM systems promises the capacity to automate many tasks previously only performed and performable by humans or to assist in ever-more complex tasks. This technology thus offers large potential benefits, including reducing tedious labour for millions as well as improving the accuracy of human decisions and actions. Applications of the technology also will open new markets for innovative and profitable businesses, such as those built on self-driving vehicles.

At the same time, however, widespread adoption of automated decision-making systems will be economically disruptive and raise complex and significant new societal challenges. These will include, but certainly not be limited to: worker displacement; machine-induced accidents; and, perhaps most fundamentally, confusion and debate over what it means to be human.

Systems built on an immature and rapidly evolving technology such as ML will have spectacular successes and dismaying failures. Especially when the technology is used in applications that affect the safety and livelihood of many people, these systems should be developed and deployed with special care. Society must set broad parameters for what uses are acceptable, how the systems should be developed, how inevitable trade-offs and conflicts will be adjudicated, and who is legally responsible for these systems and their failures.

Automated decision making is not just a scientific challenge; it is simultaneously a political, economic, technological, cultural, educational and even philosophical challenge. Because all these aspects are interconnected, it is inappropriate to focus on any one feature of the much larger picture. The computing professions and technology industries, which together are driving these advances forward, have an obligation to start a conversation among all affected disciplines and institutions whose expertise is relevant and required to fully understand these complex issues.

Now is the time to formulate appropriately nuanced, comprehensive and ethical plans for humans and our societies to thrive when computers make decisions.

Bibliography

[1] M. Murphy, "Google's AI just cracked the game that supposedly no computer could beat," Quartz, 2Day's January 2016. [Online]. Available: https://qz.com/603313/googles-ai-just-cracked-the-game-that-supposedly-no-computer-could-beat/.

[2] Google DeepMind, "AlphaGo," nd. [Online]. Available: https://deepmind.com/research/alphago/.

[3] R. Kurzweil, The Singularity is Near, Viking Press, 2005.

[4] N. Bostrom, Superintelligence: Paths, Dangers, Strategies, Oxford University Press, 2016.

[5] L. Elliott, "Millions of UK workers at risk of being replaced by robots, study says," *The Guardian*, 24 March 2017.

[6] G. C. Allen, "China's Artificial Intelligence Strategy Poses a Credible Threat to U.S. Tech Leadership," Council on Foreign Relations, 4 December 2017. [Online]. Available: https://www.cfr.org/blog/chinas-artificial-intelligence-strategy-poses-credible-threat-us-tech-leadership.

[7] ACM US Public Policy Council and ACM Europe Policy Council, "Statement on Algorithmic Transparency and Accountability," 2017.

[8] S. Kuiper, "Introduction to Multiple Regression: How Much Is Your Car Worth?," *Journal of Statistics Education*, vol. 16, no. 3, 2008.

[9] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 28 May 2015.

[10] A. Singanamalli, H. Wang and A. Madabhushi, "Cascaded Multi-view Canonical Correlation (CaMCCo) for Early Diagnosis of Alzheimer's Disease via Fusion of Clinical, Imaging and Omic Features," *Science Reports*, 15 August 2017.

[11] European Commission, "Driving and parking patterns of European car drivers," 2012.

[12] European Automobile Manufactuers Association, "Employment Trends," nd. [Online]. Available: http://www.acea.be/statistics/tag/category/employment-trends.

[13] Google Cloud Big Data and Machine Learning Blog, "An in-depth look at Google's First Tensor Processing Unit," Google, 12 May 2017. [Online]. Available: https://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu.

[14] D. Song, *AI and Security: Lessons, Challenges & Future Directions*, 2017.

[15] ACM, "Statement on Algorithmic Transparency and Accountability," 12 January 2017. [Online]. Available: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.

[16] S. Wachter, B. Mittelstadt and L. Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," *International Data Privacy Law*, vol. 7, no. 2, pp. 76-99, May 2017.

[17] S. Wachter, B. Mittelstadt and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law of Technology*, vol. Forthcomming, 2018.

[18] Commission Nationale Informatique en & Libertés, "Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificiel," 2017.

[19] B. Friedman, P. H. Kahn Jr., A. Borning and A. Huldtgren, "Value Sensitive Design and Information Systems," in *Human-Computer Interaction and Management Information Systems: Foundations Advances in Management Information Systems*, vol. 5, M.E. Sharpe, 2006, pp. 348-372.

[20] International Organization for Standardization, "ISO 26262: Road vehicles -- Functional safety," ISO, Geneva, 2011.

[21] Federal Ministry of Transport and Digital Infrastructure, "Ethics Comission: Automated and Connected Driving," 2017.

[22] N. G. Mankiw, "The Economy Is Rigged, and Other Presidential Campaign Myths," New York TImes, 6 May 2016. [Online]. Available: https://www.nytimes.com/2016/05/08/upshot/the-economy-is-rigged-and-other-presidential-campaign-myths.html?_r=0.

[23] Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, "Final Report on the accident on 1st June 2009," BEA, 2012.

[24] J. Markoff, "Google's Next Phase in Driverless Cars: No Steering Wheel or Brake Pedals," *New York TImes*, 27 May 2014.

[25] E. Pariser, The Filter Bubble: What the Internet Is Hiding from You, Penguin Press, 2011.

[26] A. Beall, "Driverless cars could let you choose who survives in a crash," *New Scientist*, 13 October 2017.

[27] R. D. Hof, "Deep Learning: With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart," *MIT Technology Review*, November 2017.

[28] M. Murphy, "Google's AI just cracked the game that supposedly no computer could beat," Quartz, January 2016. [Online]. Available: https://qz.com/603313/googles-ai-just-cracked-the-game-that-supposedly-no-computer-could-beat/.

Contact:
James Larus, james.larus@epfl.ch Chris Hankin, c.hankin@imperial.ac.uk

INFORMATICS EUROPE

acm Europe Council

euacm ACM Europe Policy Committee

DOI: 10.1145/3185595