

## Artificial Intelligence (AI)

### Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy

Dwivedi, Yogesh K.; Hughes, Laurie; Ismagilova, Elvira; Aarts, Gert; Coombs, Crispin; Crick, Tom; Duan, Yanqing; Dwivedi, Rohita; Janssen, Marijn; More Authors

#### DOI

[10.1016/j.ijinfomgt.2019.08.002](https://doi.org/10.1016/j.ijinfomgt.2019.08.002)

#### Publication date

2019

#### Document Version

Final published version

#### Published in

International Journal of Information Management

#### Citation (APA)

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Janssen, M., & More Authors (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

#### Important note

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

#### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

#### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

***Green Open Access added to TU Delft Institutional Repository***

***'You share, we take care!' - Taverne project***

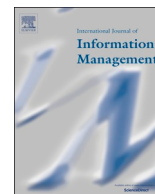
**<https://www.openaccess.nl/en/you-share-we-take-care>**

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.



Contents lists available at ScienceDirect

## International Journal of Information Management

journal homepage: [www.elsevier.com/locate/ijinfomgt](http://www.elsevier.com/locate/ijinfomgt)

## Opinion paper

## Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy

Yogesh K. Dwivedi<sup>a,\*</sup>, Laurie Hughes<sup>a</sup>, Elvira Ismagilova<sup>b</sup>, Gert Aarts<sup>c,1</sup>, Crispin Coombs<sup>d,1</sup>, Tom Crick<sup>e,1</sup>, Yanqing Duan<sup>f,1</sup>, Rohita Dwivedi<sup>g,1</sup>, John Edwards<sup>h,1</sup>, Aled Eirug<sup>i,1</sup>, Vassilis Galanos<sup>j,1</sup>, P. Vigneswara Ilavarasan<sup>k,1</sup>, Marijn Janssen<sup>l,1</sup>, Paul Jones<sup>m,1</sup>, Arpan Kumar Kar<sup>n,1</sup>, Hatice Kizgin<sup>b,1</sup>, Bianca Kronemann<sup>m,1</sup>, Banita Lal<sup>f,1</sup>, Biagio Lucini<sup>o,1</sup>, Rony Medaglia<sup>o,1</sup>, Kenneth Le Meunier-FitzHugh<sup>p,1</sup>, Leslie Caroline Le Meunier-FitzHugh<sup>p,1</sup>, Santosh Misra<sup>q,1</sup>, Emmanuel Mogaji<sup>r,1</sup>, Sujeet Kumar Sharma<sup>s,1</sup>, Jang Bahadur Singh<sup>s,1</sup>, Vishnupriya Raghavan<sup>t,1</sup>, Ramakrishnan Raman<sup>u,1</sup>, Nripendra P. Rana<sup>b,1</sup>, Spyridon Samothrakakis<sup>v,1</sup>, Jak Spencer<sup>w,1</sup>, Kuttimani Tamilmani<sup>b,1</sup>, Annie Tubadji<sup>x,1</sup>, Paul Walton<sup>y,1</sup>, Michael D. Williams<sup>z,1</sup>

<sup>a</sup> Emerging Markets Research Centre (EMaRC), School of Management, Swansea University, United Kingdom<sup>b</sup> International Business, Marketing and Branding Research Centre, School of Management, University of Bradford, Bradford, United Kingdom<sup>c</sup> Department of Physics, College of Science, Swansea University, United Kingdom<sup>d</sup> School of Business and Economics, Loughborough University, United Kingdom<sup>e</sup> School of Education, Swansea University, Swansea, United Kingdom<sup>f</sup> Business and Management Research Institute, University of Bedfordshire, United Kingdom<sup>g</sup> Prin. L.N. Welingkar Institute of Management Development & Research, Mumbai, India<sup>h</sup> Operations & Information Management Department, Aston Business School, United Kingdom<sup>i</sup> Morgan Academy, School of Management, Swansea University, United Kingdom<sup>j</sup> School of Social and Political Sciences, University of Edinburgh, United Kingdom<sup>k</sup> Department of Management Studies, Indian Institute of Technology, Delhi, India<sup>l</sup> Information and Communication Technology Section, Faculty of Technology, Policy and Management, Delft University of Technology, Netherlands<sup>m</sup> School of Management, Swansea University, United Kingdom<sup>n</sup> Department of Mathematics, Computational Foundry, Swansea University, United Kingdom<sup>o</sup> Department of Digitalization, Copenhagen Business School, Denmark<sup>p</sup> Norwich Business School, University of East Anglia, Norwich, Norfolk, United Kingdom<sup>q</sup> CEO, Commissioner of e-Governance, Government of Tamil Nadu, India<sup>r</sup> Department of Marketing, Events and Tourism, University of Greenwich, United Kingdom<sup>s</sup> Information Systems Area, Indian Institute of Management Tiruchirappalli, India<sup>t</sup> Manipal Global Education Services, Bangalore, India<sup>u</sup> Symbiosis Institute of Business Management, Pune & Symbiosis International (Deemed) University, Pune, India<sup>v</sup> Institute for Analytics and Data Science, University of Essex, United Kingdom<sup>w</sup> Urban Scale Interventions, United Kingdom<sup>x</sup> Economics Department, University of West of England, United Kingdom<sup>y</sup> Capgemini UK Ltd, United Kingdom<sup>z</sup> Swansea i-Lab (Innovation Lab), School of Management, Swansea University, United Kingdom

\* Corresponding author.

E-mail addresses: [y.k.dwivedi@swansea.ac.uk](mailto:y.k.dwivedi@swansea.ac.uk) (Y.K. Dwivedi), [d.l.hughes@swansea.ac.uk](mailto:d.l.hughes@swansea.ac.uk) (L. Hughes), [e.ismagilova@bradford.ac.uk](mailto:e.ismagilova@bradford.ac.uk) (E. Ismagilova), [g.aarts@swansea.ac.uk](mailto:g.aarts@swansea.ac.uk) (G. Aarts), [c.r.coombs@lboro.ac.uk](mailto:c.r.coombs@lboro.ac.uk) (C. Coombs), [thomas.crick@swansea.ac.uk](mailto:thomas.crick@swansea.ac.uk) (T. Crick), [yanqing.duan@beds.ac.uk](mailto:yanqing.duan@beds.ac.uk) (Y. Duan), [rohita.dwivedi@welingkar.org](mailto:rohita.dwivedi@welingkar.org) (R. Dwivedi), [j.s.edwards@aston.ac.uk](mailto:j.s.edwards@aston.ac.uk) (J. Edwards), [aled.eirug@swansea.ac.uk](mailto:aled.eirug@swansea.ac.uk) (A. Eirug), [vassilis.galanos@ed.ac.uk](mailto:vassilis.galanos@ed.ac.uk) (V. Galanos), [vignes@iitd.ac.in](mailto:vignes@iitd.ac.in) (P.V. Ilavarasan), [m.f.w.h.a.janssen@tudelft.nl](mailto:m.f.w.h.a.janssen@tudelft.nl) (M. Janssen), [w.p.jones@swansea.ac.uk](mailto:w.p.jones@swansea.ac.uk) (P. Jones), [arpan\\_kar@yahoo.co.in](mailto:arpan_kar@yahoo.co.in) (A.K. Kar), [kizgin.hatice@gmail.com](mailto:kizgin.hatice@gmail.com) (H. Kizgin), [bianca.kronemann@gmail.com](mailto:bianca.kronemann@gmail.com) (B. Kronemann), [banita.lal@beds.ac.uk](mailto:banita.lal@beds.ac.uk) (B. Lal), [b.lucini@swansea.ac.uk](mailto:b.lucini@swansea.ac.uk) (B. Lucini), [rony@cbs.dk](mailto:rony@cbs.dk) (R. Medaglia), [k.le-meunier-fitzhugh@uea.ac.uk](mailto:k.le-meunier-fitzhugh@uea.ac.uk) (K. Le Meunier-FitzHugh), [l.fitzhugh@uea.ac.uk](mailto:l.fitzhugh@uea.ac.uk) (L.C. Le Meunier-FitzHugh), [santoshmisraias@gmail.com](mailto:santoshmisraias@gmail.com) (S. Misra), [e.o.mogaji@greenwich.ac.uk](mailto:e.o.mogaji@greenwich.ac.uk) (E. Mogaji), [sujeet@iimtrichy.ac.in](mailto:sujeet@iimtrichy.ac.in) (S.K. Sharma), [jbs@iimtrichy.ac.in](mailto:jbs@iimtrichy.ac.in) (J.B. Singh), [vishnupriyaraghavan@gmail.com](mailto:vishnupriyaraghavan@gmail.com) (V. Raghavan), [director@sibmpune.edu.in](mailto:director@sibmpune.edu.in) (R. Raman), [nrananp@gmail.com](mailto:nrananp@gmail.com) (N.P. Rana), [ssamot@essex.ac.uk](mailto:ssamot@essex.ac.uk) (S. Samothrakakis), [jak@urbanscaleinterventions.com](mailto:jak@urbanscaleinterventions.com) (J. Spencer), [kuttimani.tamilmani@gmail.com](mailto:kuttimani.tamilmani@gmail.com) (K. Tamilmani), [atubadji@hotmail.com](mailto:atubadji@hotmail.com) (A. Tubadji), [paul.walton@capgemini.com](mailto:paul.walton@capgemini.com) (P. Walton), [m.d.williams@swansea.ac.uk](mailto:m.d.williams@swansea.ac.uk) (M.D. Williams).

<sup>1</sup> These authors have made equal contributions and are placed in alphabetical order.<https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

Received 11 July 2019; Received in revised form 2 August 2019; Accepted 3 August 2019

0268-4012/ © 2019 Elsevier Ltd. All rights reserved.

## ARTICLE INFO

## Keywords:

Artificial intelligence  
AI  
Cognitive computing  
Expert systems  
Machine learning  
Research agenda

## ABSTRACT

As far back as the industrial revolution, significant development in technical innovation has succeeded in transforming numerous manual tasks and processes that had been in existence for decades where humans had reached the limits of physical capacity. Artificial Intelligence (AI) offers this same transformative potential for the augmentation and potential replacement of human tasks and activities within a wide range of industrial, intellectual and social applications. The pace of change for this new AI technological age is staggering, with new breakthroughs in algorithmic machine learning and autonomous decision-making, engendering new opportunities for continued innovation. The impact of AI could be significant, with industries ranging from: finance, healthcare, manufacturing, retail, supply chain, logistics and utilities, all potentially disrupted by the onset of AI technologies. The study brings together the collective insight from a number of leading expert contributors to highlight the significant opportunities, realistic assessment of impact, challenges and potential research agenda posed by the rapid emergence of AI within a number of domains: business and management, government, public sector, and science and technology. This research offers significant and timely insight to AI technology and its impact on the future of industry and society in general, whilst recognising the societal and industrial influence on pace and direction of AI development.

## 1. Introduction

Artificial Intelligence (AI) is a concept that has been part of public discourse for decades, often depicted within science fiction films or debates on how intelligent machines will take over the world relegating the human race to a mundane servile existence in supporting the new AI order. Whilst this picture is a somewhat caricature-like depiction of AI, the reality is that artificial intelligence has arrived in the present and many of us regularly interact with the technology in our daily lives. AI technology is no longer the realm of futurologists but an integral component of the business model of many organisations and a key strategic element in the plans for many sectors of business, medicine and governments on a global scale. This transformational impact from AI has led to significant academic interest with recent studies researching the impacts and consequences of the technology rather than the performance implications of AI, which seems to have been the key research domain for a number of years.

The literature has offered various definitions of AI, each encapsulating the key concepts of non-human intelligence programmed to perform specific tasks. Russell and Norvig (2016) defined the term AI to describe systems that mimic cognitive functions generally associated with human attributes such as learning, speech and problem solving. A more detailed and perhaps elaborate characterisation was presented in Kaplan and Haenlein (2019), where the study describes AI in the context of its ability to independently interpret and learn from external data to achieve specific outcomes via flexible adaptation. The use of big data has enabled algorithms to deliver excellent performance for specific tasks (robotic vehicles, game playing, autonomous scheduling etc.) and a more pragmatic application of AI rather than the more cognitive focussed – human level AI where the complexities of human thinking and feelings have yet to be translated effectively (Hays & Efros, 2007; Russell & Norvig, 2016). The common thread amongst these definitions is the increasing capability of machines to perform specific roles and tasks currently performed by humans within the workplace and society in general.

The ability for AI to overcome some of the computationally intensive, intellectual and perhaps even creative limitations of humans, opens up new application domains within education and marketing, healthcare, finance and manufacturing with resulting impacts on productivity and performance. AI enabled systems within organisations are expanding rapidly, transforming business and manufacturing, extending their reach into what would normally be seen as exclusively human domains (Daugherty & Wilson, 2018; Miller, 2018). The era of AI systems has progressed to levels where autonomous vehicles, chatbots, autonomous planning and scheduling, gaming, translation,

medical diagnosis and even spam fighting can be performed via machine intelligence. The views of AI experts as presented in Müller and Bostrom (2016), predicted that AI systems are likely to reach overall human ability by 2075 and that some experts feel that further progress of AI towards super intelligence may be bad for humanity. Society generally is yet to fully grasp many of the ethical and economic considerations associated with AI and big data and its wider impact on human life, culture, sustainability and technological transformation (Duan, Edwards, & Dwivedi, 2019; Pappas, Mikalef, Giannakos, Krogstie, & Lekakos, 2018).

The probabilistic analysis of the economic impact of AI and automation has been assessed by the World Economic Forum (WEF), where they predict that 20% of existing UK jobs could be impacted by AI technologies. This figure is greater in emerging economies such as China and India, where the level rises to 26% due to the greater scope for technological change within the manufacturing sector. AI technologies are predicted to drive innovation and economic growth creating 133 million new jobs globally by 2022, contributing 20% of GDP within China by 2030 (WEF 2018). AI technology spending in Europe for 2019 has increased 49% over the 2018 figure to reach \$5.2 billion (IDC, 2019). Juniper Research (2019) highlighted that global spending on AI technologies within the consumer retail sector alone is predicted to reach \$12bn by 2023, a significant rise from the current figure of \$3.5bn. The research also highlighted the increasing use of AI in the form of chatbots for customer service applications, where these deployments could realise annual savings of \$439m globally by 2023, up from \$7m in 2019. Technology giants such as Amazon and Walmart have been experimenting with AI for some time, applying the technology to demand forecasting and supply chain fulfilment. Walmart's store of the future – Intelligent Retail Lab (IRL) is testing AI with analytics to trigger the need to respond when customers pick the last item and then track the store's ability to quickly restock the product. The Walmart IRL AI systems are supported by cameras and sensors installed throughout the store that transmit 1.6 TB of data per second to data centres and linked supply chain fulfilment (Forbes, 2019a). The use of AI technology within this sector can only increase as other firms respond to the competition from these market leaders.

The potential for AI has not been lost on the global superpowers with the US and China heavily focussed on the race for technology supremacy in this area. Currently this seems to be a battle that China seems to be winning with estimates of \$12 billion spending on AI in 2017 and predicted spend of up to \$20 billion by 2020. Although the Trump administration has earmarked \$2 billion for the department of Defence to spend on its AI Next project, this pales into insignificance when compared to China. Chinese academics continue to publish

significant levels of articles on AI and Chinese industry has increased the number of AI patents by 200% in recent years, significantly surpassing the US. Although Europe is still the lead academic publisher on AI related technologies, China now accounts for 25% of the global output (Shoham et al., 2018). China is determined to be the world leader in AI by 2030 (Forbes, 2019b). China's ability to aggressively implement rather than rely solely on innovation coupled with its hypercompetitive and entrepreneurial economy and business friendly governance, has driven the AI sector forward (FT, 2019).

Whilst the benefits of greater levels of AI adoption within many sectors of the global economy are felt in the context of greater efficiency, improved productivity and reliability, this picture of positive innovation is not universally welcomed globally. Estimates for work displacement due to automation, highlight that up to a third of current work activities could be impacted by 2030 (Manyika et al., 2017). Studies have analysed the impact of this significant change, developing a narrative of a changing jobs market that is predicted to focus humans further up the value chain on more creative and cognitive orientated roles in support of AI technologies (DIN & DKE, 2018; Jonsson & Svensson, 2016). However, is this particular vision of an AI future a universal one across the globe within both developed and emerging markets? The fact that AI has the capacity to replace many rules-based and repetitive tasks, means that significant numbers of jobs that traditionally would be undertaken within emerging market economies will be lost. There are benefits of AI being centred within the developed economies where new higher skilled jobs are likely to be created, but there is a potential scenario where AI could displace millions of jobs within emerging economies. This is likely to have significant impact within Asia and Africa as traditional low skilled jobs are replaced by intelligent machine thereby damaging growth and worker livelihoods within these economies (BBC, 2019). The social/economic construction of AI, its impact on humans and society from its evolution, is still being assessed. However, it is clear that there are likely to be both winners and losers and that decision makers need to be strategic in their outlook for the future.

This study brings together the collective insight from the workshop entitled "Artificial Intelligence (AI): Emerging Challenges, Opportunities, and Agenda for Research and Practice" held at the School of Management, Swansea University, UK on 13th June 2019. Contributions were received from collaborators within industry, academia and public sector to highlight the significant opportunities, challenges and potential research agenda posed by the emergence of AI

within several domains: business and management, government and public sector, science and technology. This research is presented as offering significant and timely insight to AI technology, its potential application and its impact on the future of industry and society.

The remaining sections of this article are organised as follows: Section 2 presents many of the key debates and overall themes within the literature; Section 3 details the multiple perspectives on AI technologies from the expert contributors; Section 4 presents a discussion on the key AI related topics relating to the challenges, opportunities and research agendas presented by the expert contributors. The study is concluded in Section 5.

## 2. Debate within existing literature

This section synthesises the existing AI focussed literature and elaborates on the key themes listed in Table 1 from the literature review. Studies included in this section were identified using the Scopus database, using the following combination of keywords

(TITLE ("Artificial intelligence") AND TITLE ("Advantages" OR "Benefit" OR "Opportunities" OR "Limitation" OR "Challenge" OR "Barriers" OR "Shortcoming" OR "agenda" OR "Research Direction"). This approach is similar to approach employed by existing review articles on various topics (see for example, Al-Emran, Mezhyuev, Kamaludin, & Shaalan, 2018; Dwivedi, Kapoor, & Chen, 2015a; Dwivedi & Kuljis, 2008; Hughes et al., 2019; Ismagilova, Hughes, Dwivedi, & Raman, 2019; Kapoor et al., 2018; Koivisto & Hamari, 2019; Olanrewaju, Hossain, Whiteside, & Mercieca, 2020; Senyo, Liu, & Effah, 2019; Tamilmani, Rana, Prakasam, & Dwivedi, 2019). Existing research reviewed for this article is categorised in the following major themes: AI and Decision Making; Application Domains; Data and Information; Challenges.

### 2.1. AI and decision making

Aspects of the literature have considered the use and impact of AI based systems for decision-making applications. These studies include topics such as: Algorithmic; Artificial Neural Networks; Decision Support Systems; Deep Learning; Deep Neural Networks; Expert Systems; and Learning Systems. Studies have applied artificial neural techniques to data analysis and pattern recognition problems. The research by Abbot and Marohasy (2013) examined the application of neural networks based on AI for forecasting monthly rainfall in Nebo,

**Table 1**  
Themes in AI research.

Theme	Details	Citations
AI and decision making	Artificial Neural Network Deep Learning	Abarca-Alvarez et al., 2018; Abbot & Marohasy, 2013; Baldassarre et al., 2017; Cleophas & Cleophas, 2010; Kahn, 2017; Anderson, 2019; Lassau et al., 2019; Nguyen & Shetty, 2018; Reza Tizhoosh & Pantanowitz, 2018; Stead, 2018; Thrall et al., 2018
	Algorithmic Learning Systems	Dreyer & Allen, 2018; Kahn, 2017; Risse, 2019; Stead, 2018; Varga-Szemes et al., 2018; Zandi et al., 2019; Duan et al., 2019; Glauner et al., 2017; Walton, 2018a; Walton, 2018b; Wang, Törngren, & Onori, 2015a; Wang, Li, & Leung, 2015b
Application domains	Decision Support Systems Deep Neural Networks	Abarca-Alvarez et al., 2018; Milano, O'Sullivan, & Gavanelli, 2014; Schulz & Nakamoto, 2013; Milano et al., 2014; Mitchell, 2019; Duan et al., 2019
	Robotics	Edwards, 2018; Erikson & Salzmänn-Erikson, 2016; Gupta & Kumari, 2017
	Healthcare and Informatics	Beregi et al., 2018; Cheshire, 2017; Cleophas & Cleophas, 2010; Combi, 2017; Dreyer & Allen, 2018; Gupta & Kumari, 2017; Houssami et al., 2017; Kahn, 2017; Khanna et al., 2013; Lassau et al., 2019; Nguyen & Shetty, 2018; Stead, 2018; Thesmar et al., 2019; Thrall et al., 2018; Varga-Szemes et al., 2018; Xu et al., 2019; Zandi et al., 2019
	Digital Imaging Education and Policy Manufacturing	Beregi et al., 2018; Gupta & Kumari, 2017; Kahn, 2017; Lassau et al., 2019; Nguyen & Shetty, 2018; Stead, 2018; Arltisch & Newell, 2017; Chaudhri et al., 2013; Mikhaylov et al., 2018; Nguyen, 2018; Yoon & Baek, 2016; Yoon, 2016; DIN & DKE, 2018; Haefner & Panuwatwanich, 2017; Jain & Mosier, 1992; Jonsson & Svensson, 2016; Katz, 2017; Kumar, 2017; Kusiak, 1987; Lee, 2002; Li 2018; Li et al., 2017; Löffler & Tschiesner, 2013; Makridakis, 2018; Muhuri et al., 2019; Nikolic et al., 2017; Parveen, 2018; Wang, Törngren, & Onori, 2015a; Wang, Li, & Leung, 2015b; Wang & Wang, 2016; Yang et al., 2017; Zhong et al., 2017a
Data & information	Big Data	Abarca-Alvarez et al., 2018; Beregi et al., 2018; Duan et al., 2019; Rubik & Jabs, 2018; Schulz & Nakamoto, 2013; Stead, 2018; Thrall et al., 2018; Xu et al., 2019
	Data Visualisation	Olshannikova et al., 2015; Zheng et al., 2016; Zhong et al., 2017b



Queensland, Australia. The study highlighted the benefits in combining multiple non-linear relationships using neural networks to predict rainfall patterns one month in advance. This application of AI was posited as directly contributing to the prediction of flood risk weather patterns.

Deep Learning is a term gaining traction within the literature and is associated with machine learning architectures and concepts but at a greater level and depth of neural network layers (Glauner, Meira, Valtchev, State, & Bettinger, 2017). Studies have posited the potential benefits of Deep Learning applications in areas of digital pathology and related medical applications, whilst cognisant of the limitations of this technology in terms of human reasoning and interpretation (Reza Tizhoosh & Pantanowitz, 2018; Stead, 2018). Anderson (2019) analysed the potential of combining Deep Learning technology with Electrocardiogram (ECG) applications to detect patients with reduced left ventricular ejection fraction (rLVEF). Detecting patients with rLVEF would be helpful in patients for whom echocardiography or other imaging modalities are not available or too expensive. Early diagnosis of rLVEF could directly impact patient diagnosis and mortality levels.

Studies have posited the benefits of utilising deep neural networks to improve the use of AI, however, the use of deeper networks and big datasets is unlikely to develop *meaning* in the human context, requiring further interdisciplinary research to unlock this area (Mitchell, 2019).

## 2.2. Application domains

The AI literature has identified several separate domains in which the technology can be applied: Digital Imaging, Education, Government, Healthcare, Manufacturing, Robotics and Supply Chain. Studies have analysed the impact of AI and its potential to replace humans via intelligent automation within manufacturing, supply chain, production and even the construction industry (Kusiak, 1987; Muhuri, Shukla, & Abraham, 2019; Parveen, 2018). Existing factory processes will be increasingly subject to analysis to ascertain whether they could be automated (Lee, 2002; Löffler & Tschiesner, 2013; Yang, Chen, Huang, & Li, 2017). AI centric technologies will be able to monitor and control processes in real time offering significant efficiencies over manual processes (Jain & Mosier, 1992; Zhong, Xu, Klotz, & Newman, 2017a). Organisations have posited the benefits of integrating AI technologies in the development of intelligent manufacturing and the smart factory of the future (Li, Hou, Yu, Lu, & Yang, 2017; Nikolic, Ignjatovic, Suzic, Stevanovic, & Rikalovic, 2017). The literature has

generally moved on from the somewhat dated concepts of AI based machines replacing all human workers. Studies have recognised the realistic limits of the continuing drive to automation, highlighting a more realistic *human in the loop* concept where the focus on AI is to enhance human capability, not replace it (Katz, 2017; Kumar, 2017). Humans are likely to move up the value chain to focus on design and integration related activities as part of an integrated AI, machines and human based workforce (DIN & DKE, 2018; Jonsson & Svensson, 2016; Makridakis, 2018; Wang, Törngren, & Onori, 2015a; Wang, Li, & Leung, 2015b; Wang & Wang, 2016). Manufacturing organisations are likely to use AI technologies within a production environment where intelligent machines are socially integrated within the manufacturing process, effectively functioning as co-workers for key tasks or to solve significant problems (Haeffner & Panuwatwanich, 2017).

Khanna, Sattar, and Hansen (2013) emphasised the importance of AI in healthcare, particularly in medical informatics. There is a growing requirement for new technologies that understand the complexities of hospital operations and provide the necessary productivity gains in resource usage and patient service delivery. AI has the potential to offer improved patient care and diagnosis as well as interpretation of medical imaging in areas such as radiology (Dreyer & Allen, 2018; Kahn, 2017). Screening for breast cancer (BC) and other related conditions could be more accurate and efficient using AI technology. Houssami et al.'s (2017) study analyses the use of AI for BC screening highlighting its potential in reducing false-positives and related human detection errors. The study acknowledges some of the interrelated ethical and societal trust factors but the boundaries of reliance on AI and acceptable *human in the loop* involvement is still to be developed. The application of AI and related digital technologies within public health is rapidly developing. However, collection, storage, and sharing of AI technology derived large data sets, raises ethical questions connected to governance, quality, safety, standards, privacy and data ownership (Zandi, Reis, Vayena, & Goodman, 2019). Thesmar et al. (2019) posited the benefits of utilising AI technology for insurance claims within healthcare. Claim submission, claim adjudication and fraud analysis can significantly benefit from AI use.

Education and information search is an area where the literature has identified the potential benefits of AI technology solutions. Chaudhri, Lane, Gunning, and Roschelle (2013) discussed application of AI in education to improve teacher effectiveness and student engagement. The study analysed the potential of AI within education in the context of intelligent game-based learning environments, tutoring systems and

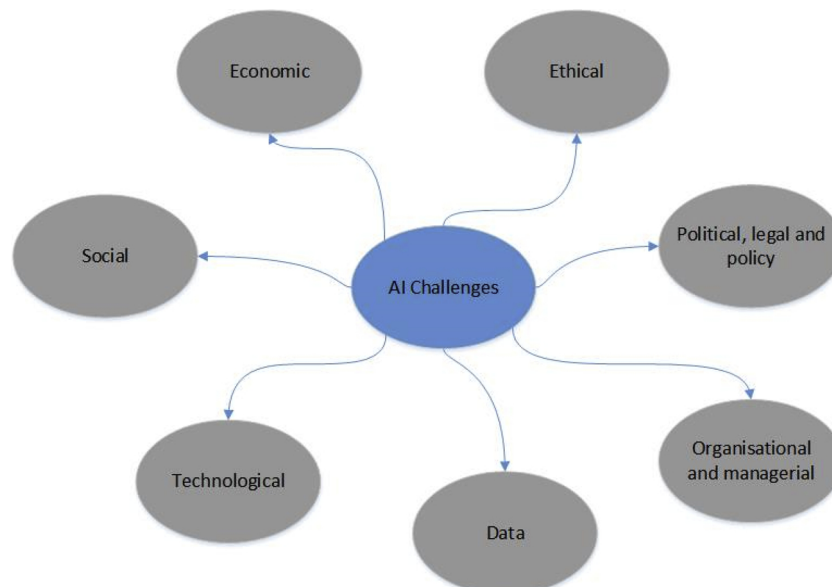


Fig. 1. AI challenges scope.

intelligent narrative technologies. The relevance of libraries in the modern technology era has received focus within the literature. [Arlitsch and Newell \(2017\)](#) discussed how AI can change library processes, staffing requirements and library users. It is important for libraries to focus on human qualities and the value add of human interaction integrated with AI to provide a richer user experience. Moreover, [Mikhailov, Esteve, and Campion \(2018\)](#) considered the use of AI capabilities from the perspective of educating the public on policy and a more effective mechanism for high uncertainty environments.

### 2.3. Data and information

The topic of big data and its integration with AI has received significant interest within the wider literature. Studies have identified the benefits of applying AI technologies to big data problems and the significant value of analytic insight and predictive capability for a number of scenarios ([Rubik & Jabs, 2018](#)). Health related studies that have analysed the impact and contribution of big data and AI arguing that these technologies can greatly support patient health based diagnosis and predictive capability ([Beregi et al., 2018](#); [Schulz & Nakamoto, 2013](#)). Big Data Analytics (BDA) develops the methodological analysis of large data structures, often categorised under the terms: volume, velocity, variety, veracity and value adding. BDA combined with AI has the potential to transform areas of manufacturing, health and business intelligence offering advanced incites within a predictive context ([Abarca-Alvarez, Campos-Sanchez, & Reinoso-Bellido, 2018](#); [Shukla, Tiwari, & Beydoun, 2018](#); [Spanaki, Gürgüç, Adams, & Mulligan, 2018](#); [Wang and Wang, 2016](#)).

Organisations are increasingly deploying data visualisation tools and methods to make sense of their big data structures. In scenarios where the limitations of human perception and cognition are taken into account, greater levels of understanding and interpretation can be gained from the analysis and presentation of data using AI technologies ([Olshannikova, Ometov, Koucheryav, & Olsson, 2015](#)). The analysis and processing of complex heterogeneous data is problematic. Organisations can extract significant value and key management information from big data via intelligent AI based visualisation tools ([Zheng, Wu, Chen, Qu, & Ni, 2016](#); [Zhong, Xu, Chen, & Huang, 2017b](#)).

### 2.4. Challenges

The implementation of AI technologies can present significant

challenges for government and organisations as the scope and depth of potential applications increases and the use of AI becomes more mainstream. These challenges are categorised in [Fig. 1](#) and discussed in this section.

[Table 2](#) lists the specific AI challenges from the literature and breakdown subtext of challenge details.

#### 2.4.1. Social challenges

The increasing use of AI is likely to challenge cultural norms and act as a potential barrier within certain sectors of the population. For example, [Xu et al. \(2019\)](#) highlighted the challenges that AI will bring to healthcare in the context of the change in interaction and patient education. This is likely to impact the patient as well as the clinician. The study highlighted the requirement for clinicians to learn to interact with AI technologies in the context of healthcare delivery and for patient education to mitigate the fear of technology for many patient demographics ([Xu et al., 2019](#)). [Theall et al. \(2018\)](#) argued that culture is one of the key barriers of AI adoption within radiology, as patients may have a reticence to interact with new technologies and systems. Social challenges have been highlighted as potential barriers to the further adoption of AI technologies. [Sun and Medaglia \(2019\)](#) identified social challenges relating to unrealistic expectations towards AI technology and insufficient knowledge on values and advantages of AI technologies. Studies have also discussed the social aspects of potential job losses due to AI technologies. This specific topic has received widespread publicity in the media and debated within numerous forums. The study by [Risse \(2019\)](#) proposed that AI creates challenges for humans that can affect the nature of work and potential influence on people's status as participants in society. Human workers are likely to progress up the value chain to focus on utilising human attributes to solve design and integration problems as part of an integrated AI and human centric workforce ([DIN & DKE, 2018](#); [Jonsson & Svensson, 2016](#); [Makridakis, 2018](#); [Wang, Törngren, & Onori, 2015a](#); [Wang, Li, & Leung, 2015b](#); [Wang & Wang, 2016](#)).

#### 2.4.2. Economic challenges

The mass introduction of AI technologies could have a significant economic impact on organisations and institutions in the context of required investment and changes to working practices. [Reza Tizhoosh and Pantanowitz \(2018\)](#) focused on the affordability of technology within the medical field arguing that AI is likely to require substantial financial investment. The study highlighted the impact on pathology

**Table 2**  
AI Challenges from the literature.

AI Challenge	Details
Social challenges	Patient/Clinician Education; Cultural barriers; Human rights; Country specific disease profiles; Unrealistic expectations towards AI technology; Country specific medical practices and insufficient knowledge on values and advantages of AI technologies.
Economic challenges	Affordability of required computational expenses; High treatment costs for patients; High cost and reduced profits for hospitals; Ethical challenges including: lack of trust towards AI based decision making and unethical use of shared data.
Data challenges	Lack of data to validate benefits of AI solutions; Quantity and quality of input data; Transparency and reproducibility; Dimensionality obstacles; Insufficient size of available data pool; Lack of data integration and continuity; Lack of standards of data collection; Format and quality; Lack of data integration and continuity and lack of standards for data collection; Format and quality.
Organisational and managerial challenges	Realism of AI; Better understanding of needs of the health systems; Organisational resistance to data sharing; Lack of in-house AI talent; Threat of replacement of human workforce; Lack of strategy for AI development; Lack of interdisciplinary talent; Threat to replacement of human workforce.
Technological and technology implementation challenges	Non-Boolean nature of diagnostic tasks; Adversarial attacks; Lack of transparency and interpretability; Design of AI systems; AI safety; Specialisation and expertise; Big data; Architecture issues and complexities in interpreting unstructured data.
Political, legal and policy challenges	Copyright issues; Governance of autonomous intelligence systems; Responsibility and accountability; privacy/safety; National security threats from foreign-owned companies collecting sensitive data, Lack of rules of accountability in the use of AI; Costly human resources still legally required to account for AI based decision; Lack of official industry standards of AI use and performance evaluation.
Ethical challenges	Responsibility and explanation of decision made by AI; processes relating to AI and human behaviour, compatibility of machine versus human value judgement, moral dilemmas and AI discrimination

laboratories where current financial pressures may be exacerbated by the additional pressures to adopt AI technologies. [Sun and Medaglia \(2019\)](#) identified several healthcare related economic challenges arguing that the introduction of AI based technologies is likely to influence the profitability of hospitals and potentially raise treatment costs for patients.

AI technologies have the potential to affect many sectors within the global economy. The McKinsey report on the economic impact of AI ([Bughin, Seong, Manyika, Chui, & Joshi, 2018](#)) develops a narrative of how organisations are likely to adopt this technology and the potential challenges for key markets during the transition. The report analyses: organisation behaviours and how they are likely to adopt AI; disruption during transition as firms experience the economic gains and losses; country specific impacts where AI could potentially widen the gap amongst emerging and developed markets as well as the rich and poor ([Bughin et al., 2018](#)).

#### 2.4.3. Data challenges

The challenges of AI and integration with big data have been discussed within several studies. There is a need for new and efficient technologies to handle the large volume, variety and velocity of big data ([Khanna et al., 2013](#)). [Xu et al. \(2019\)](#) identified data challenges of using AI in cancer genomics. The study identified the challenge in validating the benefits of AI solutions and challenges in obtaining statistically significant patient outcome data. Challenges surrounding transparency and reproducibility were also highlighted, especially in the context of acceptability relating to public perception. Challenges within computational pathology and the use of AI have been discussed in [Reza Tizhoosh and Pantanowitz \(2018\)](#). The authors highlighted the complexities of using artificial neural networks in the interpretation of imagery and the dimensionality obstacle. Whilst [Varga-Szemes, Jacobs, and Schoepf \(2018\)](#) highlighted the challenges of machine learning within a cardiac imaging context, positing a need to create a standardised format to share data across different institutions. The current position on standards and data structures can be a barrier to application of AI. [Sun and Medaglia \(2019\)](#) highlighted several data challenges surrounding the use of data and data integrity. As the transition to AI technologies matures, these challenges will need to be resolved to ensure full confidence by all stakeholders.

#### 2.4.4. Organisational and managerial challenges

The transition towards adopting AI technologies presents a number of organisational and managerial challenges that have strategic implications for firms. [Reza Tizhoosh and Pantanowitz \(2018\)](#) highlighted the significant challenges in the implementation of AI. Success relating to AI adoption is likely to be evidence based, will depend on ease of use, financial return on investment and trust. The study by [Khanna et al. \(2013\)](#) highlighted the need for AI researchers to more efficiently understand the urgent current needs of health systems and design technologies in order to address them. Current AI systems need to use more sophisticated technologies where human vs computer interaction can be improved and connected with the flow of information. Studies have highlighted that organisations face significant issues where the lack of a strategy relating to implications of AI could affect critical business areas and fail to address concerns from the human workforce ([Sun & Medaglia, 2019](#)).

#### 2.4.5. Technological and technology implementation challenges

Studies have analysed the non-boolean nature of diagnostic tasks within healthcare and the challenges of applying AI technologies to the interpretation of data and imaging. [Reza Tizhoosh and Pantanowitz \(2018\)](#) highlighted the fact that humans apply cautious language or descriptive terminology, not just binary language whereas AI based systems tend to function as a black box where the lack of transparency acts as a barrier to adoption of the technology. These points are reinforced in [Cleophas and Cleophas \(2010\)](#) and [Kahn \(2017\)](#) where the

research identified several limitations of AI for imaging and medical diagnosis, thereby impacting clinician confidence in the technology. [Cheshire \(2017\)](#) discusses the limitation of medical AI-loopthink. The term loopthink is defined as a type of implicit bias, which does not perform correct reappraisal of information or revision of an ongoing plan of action. Thus, AI would disfavour qualitative human moral principles. Weak loopthink refers to the intrinsic inability of computer intelligence to redirect executive data flow because of its fixed internal hard writing, un-editable sectors of its operating system, or unalterable lines of its programme code. Strong loopthink refers to AI suppression due to internalisation of the ethical framework.

Challenges exist around the architecture of IA systems and the need for sophisticated structures to understand human cognitive flexibility, learning speed and even moral qualities ([Baldassarre, Santucci, Cartoni, & Caligiore, 2017; Edwards, 2018](#)). [Sun and Medaglia \(2019\)](#) reviewed the technological challenges of algorithm opacity and lack of ability to read unstructured data. The [Thrall et al. \(2018\)](#) study considered the challenge of a limited pool of investigators trained in AI and radiology. This could be solved by recruiting scientists with backgrounds in AI, but also by establishing educational programmes in radiology professional services ([Nguyen & Shetty, 2018; Thrall et al., 2018](#)). [Varga-Szemes et al. \(2018\)](#) highlighted that machine learning algorithms should be created by machine learning specialists with relevant knowledge of medicine and an understanding of possible outcomes and consequences. [Mitchell \(2019\)](#) highlighted that AI systems do not yet have the essence of human intelligence. AI systems are not able to understand the situations humans experience and derive the right meaning from it. This barrier of meaning makes current AI systems vulnerable in many areas but particularly to hacker attacks titled – “adversarial examples”. In these kinds of attacks, a hacker can make specific and subtle changes to sound, image or text files, which will not have a human cognitive impact but could cause a programme to make potentially catastrophic errors. As the programmes do not understand the inputs they process and outputs they produce, they are susceptible to unexpected errors and undetectable attacks. These impacts can influence domains such as: computer vision, medical image processing, speech recognition and language processing ([Mitchell, 2019](#)).

#### 2.4.6. Political, legal and policy challenges

[Gupta and Kumari \(2017\)](#) discussed legal challenges connected to AI-responsibility when errors occur using AI systems. Another legal challenge of using AI systems can be the issue of copyrights. Current legal framework needs significant changes in order to effectively protect and incentivise human generated work ([Zatarain, 2017](#)). [Wirtz, Weyerer, and Geyer \(2019\)](#) focused on the challenges of implementing AI within government positing the requirement for a more holistic understanding of the range and impact of AI-based applications and associated challenges. The study analysed the concept of AI law and regulations to control governance including autonomous intelligence systems, responsibility and accountability as well as privacy/safety.

Studies have identified the complexities of implementing AI based systems within government and the public sector. [Sun and Medaglia \(2019\)](#) used a case study approach to analyse the challenges of applying AI within the public sector in China. The study analysed three groups of stakeholders – government policy-makers, hospital managers/doctors, and IT firm managers to identify how they perceive the challenges of AI adoption in the public sector. The study analysed the scope of changes and impact on citizens in the context of: Political, legal and policy challenges as well as national security threats from foreign-owned companies.

#### 2.4.7. Ethical challenges

Researchers have discussed the ethical dimensions of AI and implications for greater use of the technology. Individuals and organisations can exhibit a lack of trust and concerns relating to the ethical dimensions of AI systems and their use of shared data ([Sun & Medaglia,](#)



2019). The rapid pace of change and development of AI technologies increases the concerns that ethical issues are not dealt with formally. It is not clear how ethical and legal concerns especially around responsibility and analysis of decisions made by AI based systems can be solved. Adequate policies, regulations, ethical guidance and a legal framework to prevent the misuse of AI should be developed and enforced by regulators (Duan et al., 2019). Gupta and Kumari (2017) reinforces many of these points highlighting the ethical challenges relating to greater use of AI, data sharing issues and inoperability of systems. AI based systems may exhibit levels of discrimination even though the decisions made do not involve humans in the loop, highlighting the criticality of AI algorithm transparency (Bostrom & Yudkowsky, 2011).

## 2.5. Future opportunities

AI technology in all its forms is likely to see greater levels of adoption within organisations as the range of applications and levels of automation increase. Studies have estimated that by 2030, 70 per cent of businesses are likely to have adopted some form of AI technology within their business processes or factory setting (Bughin et al., 2018). Studies have posited the benefits of greater levels of adoption of AI within a range of applications, with manufacturing, healthcare and digital marketing developing significant academic interest (Juniper Research, 2018).

The factories of the future are likely to utilise AI technology extensively, as production becomes more automated and industry migrates to a more intelligent platform using AI and cyber physical systems (Wang & Wang, 2016). Within healthcare related studies, researchers have proposed new opportunities for the application of AI within medical diagnosis and pathology where mundane tasks can be automated with greater levels of speed and accuracy (Reza Tizhoosh & Pantanowitz, 2018). Through the use of human biofield technology, AI systems linked to sensors placed on and near the human body can monitor health and well-being (Rubik & Jabs, 2018). AI technologies will be able to monitor numerous life-signs parameters via Body Area Networks (BANs) where remote diagnosis requiring specialised clinical opinion and intervention will be checked by a human (Hughes, Wang, & Chen, 2012).

AI technologies have been incorporated into marketing and retail where big data analytics are used to develop personalised profiles of customers and their predicted purchasing habits. Understanding and predicting consumer demand via integrated supply chains is more critical than ever and AI technology is likely to be a critical integral element. Juniper Research (2018) predicts that demand forecasting using AI will more than treble between 2019 and 2023 and that chatbot interactions will reach 22bn in the same year from current levels of 2.6bn. The study highlights that firms are investing heavily in AI to improve trend analysis, logistics planning and stock management. AI based innovations such as the virtual mirror and visual search are set to improve the customer interaction and narrow the gap between the physical and virtual shopping experience (Juniper Research, 2018).

Researchers have argued for the more realistic future where the relationship between AI is likely to transition towards a *human in the loop* collaborative context rather than an industry-wide replacement of humans (Katz, 2017; Kumar, 2017). Stead (2018) asserts the importance of establishing a partnership where the AI machine will calculate and/or predict and humans will explain and decide on the appropriate action. Humans are likely to focus on more value add activities requiring design, analysis and interpretation based on AI processing and outputs. Future organisations are likely to focus on creating value from an integrated human and AI collaborative workforce (Jonsson & Svensson, 2016; Makridakis, 2018; Wang, Törngren, & Onori, 2015a; Wang, Li, & Leung, 2015b; Wang & Wang, 2016).

## 3. Multiple perspectives from invited contributors

This section has been structured by employing an approach adopted from Dwivedi et al. (2015b) to present consolidated yet multiple perspectives on various aspects of AI from invited expert contributors. We invited each expert to set out their contribution in up to 3–4 pages, which are compiled in this section in largely unedited form, expressed directly as they were written by the authors. Such an approach creates an inherent unevenness in the logical flow but captures the distinctive orientations of the experts and their recommendations at this critical juncture in the evolution of AI (Dwivedi et al., 2015b). The list of topics and contributors is presented in Table 3.

**Table 3**  
Invited contributor subject list.

Title of AI related topic	Author(s)
<b>Technological perspectives</b>	
Explainability and AI systems	John S. Edwards
Information Theoretic Challenges, Opportunities & Research Agenda	Paul Walton
<b>Business and management perspective</b>	
A Decision-Making Perspective	Yanqing Duan, John Edwards, Yogesh Dwivedi
AI-enabled Automation	Crispin Coombs
Labour Under Partial and Complete Automation	Spyros Samothrakis
A Generic Perspective of AI	Arpan Kar
Artificial Intelligence for Digital Marketing	Emmanuel Mogaji
Artificial Intelligence for Sales	Kenneth Le Meunier-Fitzhugh, Leslie Caroline Le Meunier-FitzHugh
Complementary Assets and Affordable-tech as Pathways for AI in the Developing World: Case of India	Vigneswara Ilavarasan
<b>Arts, humanities &amp; law perspective</b>	
People-Centred Perspectives on Artificial Intelligence	Jak Spencer
Taste, Fear and Cultural Proximity in the Demand for AI Goods and Services	Annie Tubadji
<b>Science and technology perspective</b>	
Perspectives on Artificial Intelligence in the fundamental sciences	Gert Aarts, Biagio Lucini
Science and Technology Studies –	Vassilis Galanos
<b>Government and public sector perspective</b>	
Artificial Intelligence in the public sector	Rony Medaglia
AI for SMEs and Public Sector Organisations	Sujeet Sharma and JB Singh
Public Policy Challenges of Artificial Intelligence (AI): A New Framework and Scorecard for Policy Makers and Governments	Santosh K Misra
Governance of AI and connected systems	Marijn Janssen

### 3.1. Technological perspective

#### 3.1.1. Explainability and AI systems – John S. Edwards

Explainability is the ability to explain the reasoning behind a particular decision, classification or forecast. It has become an increasingly topical issue recently in both theory and practice of AI and machine learning systems.

**3.1.1.1. Challenges.** Explainability has been an issue ever since the earliest days of AI use in business in the 1980s. This accounted for much of the early success of rule-based expert systems, where explanations were straightforward to construct, compared to frame-based systems, where explanations were more difficult, and neural networks, where they were impossible. At their inception, neural networks were unable to give explanations except in terms of weightings with little real-world relevance. As a result, they were often referred to as “black box” systems. More recently, so-called deep learning systems (typically neural networks with more than one hidden layer) make the task of explanation even more difficult.

The implied “gold standard” has been that when a person makes a decision, they can be asked to give an explanation, but this human explanation process is a more complex one than is usually recognised in the AI literature, as indicated by Miller (2019). Even if a human explanation is given that appears valid, is it accurate? Face-to-face job interviews are notorious for the risk of being decided on factors (such as how the interviewee walks across the room) other than the ones the panel members think they are using. This is related to the difficulty of making tacit knowledge explicit.

There is also a difference between the “how” explanations that are useful for AI system developers and the “why” explanations that are most helpful to end-users. Preece (2018) describes how this too was recognised in the earliest days of expert systems such as MYCIN. Nevertheless, some of the recent AI literature seems unaware of this; it is perhaps significant that the machine learning literature tends to use the term interpretability rather than explainability. There are, however, many exceptions such as Adadi and Berrada (2018), who identify four reasons for explanation: to justify, to control, to improve and to discover.

An important change in context is that governments are now introducing guidelines for the use of any type of automated decision-making systems, not just AI systems. For example, the European Union's General Data Protection Regulation (GDPR) Article 22 states “The data subject shall have the right not to be subject to a decision based solely on automated processing”, and the associated Recital 71 gives the data subject “the right...to obtain an explanation of the decision reached after such assessment and to challenge the decision”. Similarly, the UK government has introduced a code of conduct for the use of “data-driven technology” in health and social care (Anonymous, 2018). In regulated industries, existing provisions about decision-making, such as outlawing “red-lining” in evaluating mortgage or loan applications, which were first enshrined in law in the United States (US) as far back as the 1960s, also apply to AI systems.

**3.1.1.2. Opportunities.** People like explanations, even when they are not really necessary. It is not a major disaster if Netflix® recommends a film I don't like to me, but even there a simple explanation like “because you watched < name of film/TV programme > ” is added. Unfortunately, at the time of writing, it doesn't matter whether I watched that other film/TV programme all the way through or gave up after five minutes. There is plenty of scope for improving such simple explanations. More importantly, work here would give a foundation for understanding what really makes a good explanation for an automated decision, and this understanding should be transferable to systems which need a much higher level of responsibility, such as safety-critical systems, medical diagnosis systems or crime detection systems.

Alternatively, a good explanation for an automated decision may

not need to be judged on the same criteria that would be used for a human decision, even in a similar domain. People are good at recognising faces and other types of image, but most of us do not know how we do it, and so cannot give a useful explanation. Research into machine learning-based image recognition is relatively well advanced. The work of researchers at IBM and MIT on understanding the reasoning of generative adversarial networks (GANs) for image recognition suggests that “to some degree, GANs are organising knowledge and information in ways that are logical to humans” (Dickson, 2019). For example, one neuron in the network corresponds to the concept “tree”. This line of study may even help us to understand how we humans do some tasks.

Contrary to both of these views, London (2019) argues that in medical diagnosis and treatment, explainability is less important than accuracy. London argues that human medical decision-making is not so different from a black box approach, in that there is often no agreed underlying causal model: “Large parts of medical practice frequently reflect a mixture of empirical findings and inherited clinical culture.” (p.17) The outputs from a deep learning black box approach should therefore simply be judged in the same way, using clinical trials and evidence-based practice, and research should concentrate on striving for accuracy.

Lastly, advances in data visualisation techniques and technology offer the prospect of completely different approaches to the traditional “explanation in words”.

**3.1.1.3. Research agenda.** We offer suggestions for research in five linked areas.

- Can explanations from a single central approach be tailored to different classes of explainee? Explanation approaches are typically divided into transparency and post hoc interpretation (see e.g. Preece, 2018), the former being more suitable for “how” explanations, the latter for “why”. Is it possible to tailor explanations from a single central approach to different classes of explainee (developers, end-users, domain experts...)? For example, a visualisation approach for end-users that would allow drill-down for more knowledgeable explainees?
- What sort of explanation best demonstrates compliance with statute/regulation? For example, how specific does it have to be? UK train travellers often hear “this service is delayed because of delays to a previous service”, which is a logically valid but completely useless explanation. Do there need to be different requirements for different industry sectors? What form should the explanation take – words, pictures, probabilities? The latter links to the next point.
- Understanding the validity and acceptability of using probabilities in AI explanation. It is well-known that many people are poor at dealing with probabilities (Tversky & Kahneman, 1983). Are explanations from AI systems in terms of probabilities acceptable? This is widely used in the healthcare sector already, but it is not clear how well understood even the existing explanations are, especially in the light of the comments by London mentioned in the previous section.
- Improving explanations of all decisions, not just automated ones. Can post hoc approaches like the IBM/MIT work on GANs produce better explanations of not only automated decisions, but also those made by humans?
- Investigating the perceived trade-off between transparency and system performance. It is generally accepted that there is an inverse relationship between performance/accuracy and explainability for an AI system, and hence a trade-off that needs to be made. For example, Niel Nickolaisen, vice president and CTO at human resource consulting company O.C. Tanner observed: “I agree that there needs to be some transparency into the algorithms, but does that weaken the capabilities of the [machine learning] to test different models and create the ensemble that best links cause and effect?” (Holak,

2018). Does this trade-off have to be the case? Could a radical approach to explanation be an outlier to the trade-off curve?

### 3.1.2. Information theoretic challenges, opportunities & research agenda – Paul Walton

AI is introducing new ways in which organisations can process information. Therefore, it is important to consider AI in the context of the limitations in this processing predicted by information theory (Walton, 2018a; Walton, 2018b) and consequent implications for the implementation, adoption and use of AI.

**3.1.2.1. Challenges.** The implementation of AI is different from traditional technology implementation and introduces a new set of challenges. These challenges are driven by a combination of the following factors: the changing nature of the business environment, the nature of AI and machine learning (ML) themselves, and underlying information theory limitations that apply to all information processing but in specific ways to AI/ML.

Entities (like people, animals, organisations or computer systems) that interact with their environments are subject to information-related selection pressures that drive trade-offs between information measures—the pace and friction of information processing and the quality of the information produced (Walton, 2014; Walton, 2015a; Walton, 2015b). These selection pressures occur differently in alternate environments, so information ecosystems have developed with alternative ways of exchanging information (e.g. languages, jargon, computer protocols). Ecosystems have their own conventions for information processes and measures driven by the trade-offs. The conventions of different ecosystems mean that each approaches information processing from a different perspective and has its own set of inherent limitations with respect to information—good enough for the selection pressures of the ecosystem but not necessarily more widely (Walton, 2017).

This concept of a viewpoint applies at different levels—to an organisation, its departments, computer systems, people and beyond. Humans, for example, have different forms of perception and inference (Mercier & Sperber, 2017) that operate in different ways and from different viewpoints. Usually the brain manages to create an integrated overall picture from these but beneath the surface impression there are gaps that magic, for example, exploits (Macknik & Martinez-Conde, 2011). Are similar gaps possible for the implementation of AI in organisations? AI is becoming pervasive (since it is just another set of computing techniques that any developer or product supplier can use) and is increasingly being included in components as diverse as smart assistants, modules for enterprise products, widely available cloud libraries and bespoke data-science-driven applications. In addition, it is being applied to numerous different business use cases. Critically, in subsets of these components, the data science may be handled independently, for example by product or cloud suppliers, with different viewpoints. So the following question arises: *how can the inferences delivered by different AI components be integrated coherently when they may be based on different data, and subject to different ecosystem conventions (and the associated quality differences)* (Walton, 2018a; Walton, 2018b)?

This question can be retitled as the discrimination problem (Walton, 2018a; Walton, 2018b)—what quality of data and inference is required to discriminate reliably between alternatives that lead to significantly different actions and outcomes? For individual AI components this translates into an analysis of the risk and tolerance associated with false-positives and false negatives. *But when multiple AI components rely on different data and ecosystem conventions, under what circumstances can organisations integrate them to enable successful discrimination? Under what circumstances will AI be sufficient and when will it need to be supported by causal reasoning or simulation* (Pearl & MacKenzie, 2018)?

For many business challenges—the management of compliance regulations is an obvious example—rationale is important; the reason for an answer is as important as the answer itself. However, deep learning does not support this well even in the case of single AI

components (although work is underway (Foy, 2018)). *In the case of multiple AI components, how can an organisation overcome this transparency challenge?*

This is one example of a deeper underlying problem, that of ecosystem boundaries. One type of ecosystem boundary, between AI and humans, is especially important (Fry, 2018). As AI tackles more complex topics the ability to exchange complex information successfully between AI components and people will become ever more important, leading to the question: *how can an organisation ensure that AI and people can work together successfully?*

The potential biases associated with AI are well known (DeBrusk, 2018). They highlight a wider question: *how can an organisation assure the outcome of integrated AI components against a range of organisational requirements, not just for individual interactions but over multiple interactions?*

These questions introduce the first two levels of fitness. The concept of fitness within an ecosystem (which measures how well an information processing entity fits its environment (Ford, 2017; Walton, 2018a; Walton, 2018b)) breaks down into three levels:

1. narrow fitness: the ability to achieve favourable outcomes in a single interaction;
2. broad fitness: the ability to achieve favourable outcomes over multiple interactions, potentially of different types (this is the level that reveals bias and, more generally, ethical and social issues);
3. adaptiveness: the ability to achieve favourable outcomes when the environment (determined by the frequency and nature of interactions) changes.

Note that there is a tension between these—an excessive focus on one can diminish the ability to achieve the others.

Organisations have a set of internal selection pressures (created from budget processes, culture, performance management, organisation design and others) that are supposed to make the organisation fit for its environment. However, they do not always align effectively with each other or the environment. In an era of disruption, especially, organisations need internal selection pressures with a different balance between the levels of fitness. For example, the difficulty that organisations have with transformation (Capgemini Report, 2018a) shows that organisations may have insufficient internal selection pressures to support adaptiveness (often resulting in high levels of friction associated with change) and consequently struggle to keep pace with changes in their environment. Current business pressures relate directly to this—the need for greater organisational responsiveness (Capgemini Report, 2018a) means that adaptiveness is more important than before. In addition, the digital and AI revolution means that end-to-end information quality (Westerman, Bonnet, & McAfee, 2014) is increasingly important. Since machine learning is about learning, this, in itself, poses a question: *how can an organisation use AI to learn about changes in its environment and then make the required changes quickly and reliably* (Walton, 2018a; Walton, 2018b)?

**3.1.2.2. Opportunities.** The opportunities for AI are numerous. As the authors say, with respect to AI (Capgemini Report, 2018b): “Almost any existing or new application can deliver more value by augmenting it with a touch of ‘smart.’” We can think of the opportunities in several categories:

- *The organisational environment:* making sense of the torrent of data available to understand opportunities (customer needs, attitudes and preferences, their specific and increasingly real-time context) and threats (including security threats, reputational threats and fraud) and take appropriate action;
- *Operations:* making sense of the data from operations, partners and the supply chain to understand status, predict and manage incidents and failure and improve efficiency and reliability;

- **Interaction:** using the capabilities of natural language processing and other sensing capabilities to interact with people (including employees, service users and customers);
- **Case management automation:** understanding what cases can be routinely automated and what cases need specialist intervention (and when);
- **Governance:** improving the quality of information available to support (automated or human) decisions (Kahneman points out (Kahneman, 2011) that, without conscious intervention, people are “radically insensitive to both the quality and quantity of information that gives rise to impressions and intuitions”);
- **Adaptiveness:** helping an organisation to improve its response to changes in the environment by, for example, re-learning business rules.

More generally, AI can assist organisations to develop both operational and strategic situation awareness and the ability to link that awareness through to action increasingly quickly, efficiently and effectively.

**3.1.2.3. Research Agenda.** The following paragraphs itemise a set of research questions relating to the challenges and opportunities outlined above.

- **Data:** how should organisations structure their business and technology architectures to support data engineering (and its links with IoT, digital twins and other technology trends) and data governance to support multiple AI components with different ecosystem conventions? How can they ensure that the quality of the data is sufficient to support the required analysis?
- **Discrimination:** under what circumstances and to what extent can organisations rely sufficiently on the discrimination provided by sets of integrated AI components based on different data and ecosystem conventions?
- **Assurance:** what capabilities, controls and mechanisms do organisations require to implement to understand and assure sufficiently the risks (for each level of fitness) associated with implementing single, multiple and integrated AI components?
- **Transparency:** under what circumstances do organisations require transparency of reasoning and how can this be delivered when AI components are integrated?
- **Adaptiveness:** how can AI contribute to improving the adaptiveness of organisations and how can organisations derive the appropriate balance between the different levels of fitness using AI components?
- **Working together:** how can AI be designed so that complex information can be exchanged reliably between AI and humans—how can they work together effectively?
- **Internal selection pressures:** how can AI support the development of internal selection pressures that can support the right balance between the different levels of fitness?
- **Inference approach:** for which business use cases will ML be sufficient (assuming availability of the right data) and for which will it need to be supported by different forms of causal reasoning or simulation?

## 3.2. Business and management perspective

**3.2.1. A decision-making perspective – Yanqing Duan, John Edwards, Yogesh Dwivedi**

**3.2.1.1. Challenges.** The earliest development of AI was the construction of an intelligent machine that could mimic human decision making for playing chess. Since then, using AI in decision making has been one of the most important applications in AI history. The roles of AI in decision making have been classified in various ways. Broadly speaking, AI systems can be used either to support/assist the human decision makers, or to replace them (Edwards, Duan, & Robins, 2000). More specifically, the early publication by Bader, Edwards,

Harris-Jones, and Hannaford (1988) identified six roles for knowledge based systems: Assistant, critic, second opinion, expert consultant, tutor, and automaton. As the current advancement in AI technology enables researchers to create more advanced machines, it is possible for AI to undertake more complex tasks that require cognitive capabilities which previously seemed impossible, such as making tacit judgements, sensing emotion and driving processes (Mahroof, 2019). As a result, an increasing number of jobs are autonomously performed by AI systems without human control and supervision (Zlotowski, Yogeewaran, & Bartneck, 2017). There are many reports on the benefits of AI for decision making because AI is believed to be able to reach improved decisions, to boost our analytical skills and decision-making abilities, and heighten creativity (Wilson & Daugherty, 2018). However, “with the resurgence of AI, a new human-machine symbiosis is on the horizon and a question remains: How can humans and new AIs be complementary in organisational decision making?” (Jarrahi, 2018 p. 579). Miller (2018) argues the imperative of a new human-machine symbiosis and calls for the rethink of “how humans and machines need to work symbiotically to augment and enhance each other’s capabilities.” (page 2). For example, what would be the implications of using AI for future business executives in making strategic decisions?

**3.2.1.2. Opportunities and research agenda.** To advance our knowledge and understanding on the new generation of AI systems for decision making, Duan, Edwards, and Dwivedi (2019) propose twelve research propositions in terms of conceptual and theoretical development, AI technology-human interaction, and AI implementation. Based on Duan et al. (2019)’s comprehensive review and discussion, this section provides the following specific research areas on the emerging challenges and research agenda of AI from a decision making perspective.

**Re-defining and explaining the role of AI for decision making:** Will AI be mostly accepted by human decision makers as a decision support/augmentation tool rather than as the automation of decision making to replace them? AI can play multiple roles in decision making, but there are contradictory views in the current debate on the role of the new generation AI.

Many previous studies have examined the roles of AI before the era of big data. However, considering the superpower of the new generation AI and the overwhelmingly mixed views and debate on the new role of AI in decision making, it is imperative that the role of AI should be revisited and redefined. Some argue that AI should be used to augment the human judgement rather than automation (Miller, 2018; Wilson & Daugherty, 2018) and “AI systems should be designed with the intention of augmenting, not replacing, human contributions” (Jarrahi, 2018, p. 584), but this assertion should be further supported with rigorous research and investigation with empirical evidence on how and why AI is best at providing augmentation in supporting human judgement rather than decision automation. Wilson and Daugherty (2018) argue that companies that deploy AI mainly to displace employees will see only short-term productivity gains. What is the evidence for this claim? If this is true, why and how will using AI for replacing employees not deliver the long-term gains and how can this shortcoming be overcome?

**Measuring and justifying the impact of AI on decision making performance:** How can you measure the impact of AI on human decision-making performance in a new human-machine symbiosis. Measuring the benefit of AI and its impact can be very difficult, but possible. There is a need to develop and test theoretically sound and practically feasible AI impact indicators to measure its benefits.

To address this issue, researchers need to clearly understand the role of AI in decision making process. For example, if it is in a decision support role, what is the most appropriate way to measure the AI’s impact on the human decision makers’ performance? Wilson and Daugherty (2018) claim that companies can benefit from optimising “collaboration between humans and artificial intelligence” and develop



employees' "fusion skills" that enable them to work effectively at the human-machine interface, but how can these benefits be directly measured?

**Developing and testing System design criteria for supporting decision making:** *What are the principal design criteria where AI is used within decision making in different roles? As the impact of AI in decision making will be realised via the human users, the ergonomic design of AI systems is important for their success. However, the ergonomic issues may be different between supporting, augmenting, replacing, or automating systems.*

As the effectiveness of AI systems for decision making can only be realised through its acceptance and use by the end users (Edwards et al., 2000), the system design criteria for AI based systems has been an issue since the early applications of AI. Based on our understanding of the roles of AI, whether for supporting, augmenting, replacing, or automating decision making, IS researchers need to propose the design criteria from technology-human interaction perspective for system developers to create ideal AI systems for human decision makers. For example, what are the ergonomic design issues for developing AI systems that are suitable for decision making?

**Refining and improving AI system performance while in use by decision makers:** *Can AI systems' performance for decision-making be refined and improved while the systems are in use by decision makers?*

AI can augment human decision-making, but human efforts are also required to augment AI. The unique strength of human intelligence is its ability to learn and adapt to new environment and challenges. Refining and improving performance through continuing learning has been a challenge for advancing AI until the recent advances in deep learning and Big data. Deep learning, as a subset of machine learning, has been one of the essential enablers for the renewed AI success. Can AI systems be refined and improved by deep learning while they are in use by decision makers? This question needs to be addressed by further research.

**Understanding the critical factors affecting AI's success in decision making:** *What are the critical factors that will significantly affect AI's success for decision making.*

While technology advancement may have no limit, its applications may encounter bottlenecks and unprecedented barriers. Factors affecting the use, impact, success and failure of information systems have been studied extensively (Dwivedi et al., 2015b; Dwivedi et al., 2017; Dwivedi, Rana, Jeyaraj, Clement, & Williams, 2019; Hughes, Dwivedi, Rana, & Simintiras, 2016; Hughes, Dwivedi, & Rana, 2017). There has been some work on critical success factors for implementing data mining systems (Bole, Popović, Žabkar, Papa, & Jaklič, 2015), but there is a lack of research on identifying the critical success factors affecting the current use of AI and its impact in the era of Big data.

**Understanding the relationship between culture and the use of AI in decision making:** *Does culture play any significant role in AI's success in decision making? It is believed that the acceptance of AI for decision making can be affected by different cultures and personal values. By contrast, the acceptance and successful application of AI for decision making may result in a change of culture in organisations and in individual decision-making behaviour.*

Culture has been recognised as an important influential factor in technology acceptance by many previous studies. Does culture, such as national or organisational culture, and personal and religious values, also play a critical role in acceptance/adoption and use of AI applications? For example, Gerbert, Reeves, Ransbotham, Kiron, and Spira (2018) examine "Why Chinese companies approach AI differently". Liu et al. (2018) also find a significant influence of both organisational and Chinese national culture on knowledge management. If culture does play a role, how, why and to what extent does it affect the AI success? Will the wide use of AI for supporting and automating human decision-making change culture? This is an area that has not been well explored so far, thus requiring further investigation.

**Theorising the use of AI and its impact on decision making:** *Why, how and to what extent is AI being used in and making impact on*

*organisational decision making? To address this question, it is necessary to theorise the use of AI and its impact on decision-making. Therefore, an integrated conceptual framework is needed to provide a systematic understanding of AI in decision-making.*

With the rapid increase in AI applications, many claims are made by AI developers and large corporates about its use and substantial benefits and impact. For example, according to Davenport and Ronanki (2018), a survey of 250 executives who are familiar with their companies' use of cognitive technology (a term Davenport and Ronanki explain as "next-generation AI") shows that three-quarters of them "believe that AI will substantially transform their companies within three years" (p. 110). As most similar claims are not substantiated by measurable empirical evidence and rigorous academic research, it is difficult to know how, why and to what extent AI systems are being used and impacting on individual and organisational decision-making performance and transforming organisations. This raises an opportunity for IS researchers to develop appropriate theoretical justifications on the use and impact of AI for decision making through the appropriate theoretical lens.

**3.2.2. Exploiting AI-enabled automation: challenges for organisational leaders – Crispin Coombs**

**3.2.2.1. Challenges.** Advances in AI technologies have seen a step change over the last 10 years. One consequence of these developments is the creation of new opportunities to automate existing work tasks. Automation can be defined as the execution by machine, usually a computer, of a function previously carried out by a human (Parasuraman, 1997). AI-enabled automation technologies can manage and analyse vast amounts of data, propose recommended courses of action and enact these decisions. These technologies are also able to improve their decision accuracy over time, thereby becoming increasingly more valuable to Organisations (Tarafdar, Beath, & Ross, 2017). Such 'intelligent' capabilities have enabled AI to be applied in repetitive and routine knowledge work, such as improving stock market timing and portfolio creation (Hilovská & Koncz, 2012) or identifying firms that are at most risk of bankruptcy (Chaudhuri & De, 2011). While much of the recent rhetoric assumes full automation of job roles (e.g. Frey & Osborne, 2017) a notable feature of many AI applications is the continuing need for a humans to work alongside the automation technology. Human workers are needed to either assess and confirm AI decision recommendations, enact the AI recommended course of action, or provide backup support should the AI-enabled automation produce errors or fail. This has led scholars to argue that AI-enabled automation will augment the work of humans, rather than enable wholesale substitution (Davenport & Kirby, 2016). Thus, understanding how humans work alongside AI-enabled automation will be critical to deliver the anticipated benefits of automation.

The Human Factors, Ergonomics and Safety Engineering literature has an established stream of research that examines the impact of traditional automation technologies on human workers. This literature provides an important starting point for Information Systems scholars wishing to investigate these issues (Markus, 2017). This literature suggests that to maximise the benefits of AI-enabled automation, organisational leaders are likely to be faced with four major challenges: (i) how to select tasks for automation; (ii) how to select the level of automation for each task; (iii) how to manage the impact of AI-enabled automation on human performance; and (iv) how to manage AI-enabled automation errors. Each of these challenges is briefly discussed below.

First, when considering tasks for AI-enabled automation Organisational leaders need to recognise that work tasks can be subdivided into specific stages. For example, information processing tasks comprise of (1) information acquisition; (2) information analysis; (3) decision and action selection; and (4) action implementation (Parasuraman & Wickens, 2008). AI-enabled automation may be applied to each individual stage or across all stages. Thus, Organisational leaders need to consider whether AI-enabled automation can be applied



to all the functional stages of a work task and whether this is desirable for the business process.

Second, Organisational leaders need to appreciate that each work task stage may have a different level of automation applied. The level of automation may range from Level 1, manual control where the computer offers no assistance to Level 8, autonomous control stage where the computer does everything without human notification (Vagia, Transeth, & Fjordingen, 2016). Thus, as well as selecting the appropriate work task stage to automate, Organisational leaders also need to decide how much decision-making control is given to the AI-enabled automation, and to what extent a human is kept in the loop.

Third, the level of automation selected for functional task stages may have impacts on human worker performance. These impacts may be positive or negative. For example, higher levels of automation may reduce operators' workload and achieve improved results. However, higher levels of automation may also reduce the situation awareness of the worker, and increase a tendency to overly rely on automation technology (Onnasch, Wickens, Li, & Manzey, 2014). Thus, Organisational leaders need to understand the factors that may influence human worker performance when working alongside AI-enabled automation.

Fourth, even the most reliable technological systems are likely to fail at some point and when this occurs, humans must engage in error management. Highly automated systems that do not require frequent intervention are hard for humans to keep attention to (Parasuraman & Manzey, 2010). This creates an automation paradox, where factors that positively influence performance when the automation is working well may undermine performance when the automation fails (Onnasch et al., 2014). Leaving the human out of the loop can be problematic because it leads to considerable human performance impairment if the automation fails. Thus, Organisational leaders need to understand how AI-enabled automation errors or failures can be managed. These four challenges present several opportunities for Information Systems scholars.

**3.2.2.2. Opportunities.** There has been considerable research that has examined the types of automation that may be applied to different tasks and stages of tasks (McBride, Rogers, & Fisk, 2014; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Parasuraman, 2005). These studies reveal that adopting strategies of automating tasks that machines do best and leaving the residual tasks to human workers is likely to have negative impacts on performance (Parasuraman, 1997). To assist leaders develop more sophisticated automation strategies several quantitative and qualitative models have been developed (Parasuraman, 2000). While these models provide a valuable starting point, further research is required to explore how quantitative and qualitative models may be combined to provide richer insights regarding appropriate and desirable AI-enabled automation of tasks.

The literature also indicates that selecting appropriate levels of automation is a complex process that is highly contingent on a wide range of factors (Parasuraman, 1997). The level of automation selected may be dependent on person factors (e.g. the complacency potential of the human operator, the automation training provided, and the knowledge of automation held by the worker), task factors (e.g. the consequence of automation error, the cost of verification, and lines of accountability), or cognitive load factors (e.g. associated impacts of automation on mental workload or situation awareness) (McBride et al., 2014). Thus, broad recommendations of 'medium' automation adoption levels are likely to be unhelpful as they do not take sufficient account of these important contingencies (Onnasch et al., 2014). Given the capabilities of current and projected AI-enabled automation technologies, more empirical research is needed to examine the conditions that influence the level of automation applied to task stages.

Existing research has shown that when working with automation technologies human performance may be influenced by automation complacency and automation bias. Automation complacency is defined as the poorer detection of system malfunctions under automation compared with under manual control (Parasuraman & Manzey, 2010).

For example, human operators (e.g. pilots, air traffic controllers) not conducting enough checks of system state and assuming "all is well" when in fact a dangerous condition is developing that leads to an accident. Automation bias has been defined as people using the outcome of a decision aid as a heuristic replacement for vigilant information seeking and processing (Mosier & Skitka, 1996). It may occur because the automatically generated cues are often very salient and draw the human's attention and because humans tend to ascribe greater power and authority to automated aids than to other sources of advice. Although, automation complacency and bias can speed up decision making when recommendations are correct, when the automation technology provides the incorrect recommendations it can lead to omission errors (the human does not respond to a critical situation) and commission errors (the human follows the recommendation of the automation, even though it is incorrect) (McBride et al., 2014; Parasuraman & Manzey, 2010). This presents an important research opportunity to explore and understand the factors that influence over reliance on automation and how to counter them, especially if the human is retained as the backup to AI-enabled automation.

The management of automation errors may also be influenced by automation complacency and automation bias. Automation complacency among human operators increases with higher levels of automation and higher automation reliability (McBride et al., 2014). This increase is because human workers are less aware of changes in the environment or states when the change is made by an agent other than themselves (human or automation). The risk of negative consequences associated with AI-enabled automation errors or failure increases with increasing levels of automation and in the latter functional stages of information processing (Sebok & Wickens, 2017). Thus, further research is required to investigate features that can mitigate the loss in performance, in circumstances of error or failure, with higher degrees of AI-enabled automation.

**3.2.2.3. Research agenda.** In order to address these research challenges and opportunities for understanding how humans work alongside AI-enabled automation several research priorities are proposed. First, further empirical research is needed to investigate how are decisions made regarding the work task to automate and level of AI-enabled automation to apply. As automation becomes increasingly intelligent through the application of AI the range and types of task that it may apply to are likely to grow. Although qualitative and quantitative models have been developed to conceptualise this decision-making process, they have been developed from studies that examine traditional automation technologies and adopt a functional task perspective. Although these studies help to explain how existing tasks or activities may be automated, they may not adequately explore new ways of undertaking business processes or the development of radical new business models. Thus, further research is needed to investigate how strategic applications of AI-enabled automation may redesign or create new business processes and how the role of the human worker may evolve alongside these developments, the new job roles that may be created and the skills required to undertake these roles. These studies should combine qualitative and quantitative models of automation selection and account for contingency factors such as person characteristics, task characteristics and the associated cognitive factors that may influence the level of automation applied.

A second research priority is to understand the factors that may influence human over reliance on AI-enabled automation and how they can be countered. As AI-enabled automation becomes common and reliable there is an increasing risk that humans will privilege AI recommendations and decisions over their own judgements and suffer from reduced situation awareness. While reduced situation awareness has critical implications in transport and health contexts, poor situation awareness may also bring significant risks in many business and management settings, such as imperfect AI-enabled automated decisions leading to stock market crashes or firm bankruptcy. Research that

studies the auditory and visual cuing of automated system performance, as opposed to relying on alerts when errors occur, or the system fails, would be valuable to address this research priority (Hancock et al., 2013). Research could also explore how AI could be used to design 'likely' alarms (Parasuraman, 1997) rather than relying on alarms to be definite warnings of dangerous situations, or apply adaptive automation that can vary the level of automation applied in real time (Hancock et al., 2013) to help to reduce loss of situation awareness due to out-of-the-loop unfamiliarity.

A third, related research priority concerns the use of human workers in failsafe capacities to protect against AI-enabled automation error or failure. Human workers are likely to find it difficult to undertake this failsafe role effectively in situations of high AI-enabled automation reliability and when they have few opportunities to practice performing the task. Thus, further research is needed to investigate how human workers can be trained and supported to continue to be able to effectively monitor and respond to AI-enabled automation errors and failures. For example, research could investigate how human workers can be trained to "expect the unexpected" as well as training in understanding of AI-enabled automation logic. Further, as higher levels of automation become more pervasive, AI-enabled automation will become more challenging to manage in situations of automation error or failure. Thus, further research is needed to explore how AI-enabled automated decision making can be made sufficiently transparent for a human to diagnose error creating faults. This is a critical research priority because understanding 'what happens when it goes wrong' is a key factor for Organisations wishing to increase their level of automated decision making.

Information Systems scholars have a critical role to play in shaping the agenda of how AI is applied in organisations and society in the future. It is hoped that this research agenda will be useful to scholars and contribute to an enhanced understanding of how leaders may exploit AI-enabled automation to deliver benefits for their organisations and society.

### 3.2.3. Labour under partial and complete automation – Spyros Samothrakis

One of the most significant facts in the history of labour is the universality of the belief that automation is going to reduce aggregate labour hours. On the more "progressive" side, thinkers from the whole of the political spectrum, including Keynes (2010), Nixon (Blair, 1956) and Stalin (Stalin, 1952) were adamant that through a combination of policy and technological automation, we would see a drastic reduction of working hours. For those with a more pessimistic bend, more automation meant increased unemployment; the beginning of this idea probably goes back to the luddites (David, 2015), who actively tried to remove machines from the production process as a means to preserve jobs. This belief (as part of a generalised fear of technological unemployment) has been re-iterated multiple times.

Unemployment seems to follow cyclical business patterns, with almost no scholar making a case for technological causes. This does not mean that certain professions will not disappear (Frey & Osborne, 2017), but that economy is on a permanent reconfiguration state. Working hours did gradually decrease until the 1970s; from that point onwards working hours have either stayed stable or increased (especially in the US), while extreme working hours have increased (Burger, 2015). The contradiction here is apparent – technological development has been astounding for the last 40 years, but the need for labour has not diminished.

**3.2.3.1. Outcomes after the limit case.** If we are to take the idea of work automation to its limits (but on the same time assuming that through control or technological inability, god-like AIs never materialise), humanity's ability to produce might only be constrained by the availability of raw resources; human labour will no longer be needed to supplement machine labour. In this scenario machines become non-conscious slaves, with no further human involvement in production,

even at planning or at discovery level. How close is this idealised limit to a possible real limit will play a crucial role in future societal developments, in conjunction with societal organisation. The idea of a humanity not needed for production has been maintained in various publications (Joy, 2000), but a thorough discussion is provided in Frase (2016). The argument is simple – if technological trends are to reach their limit, we can discern four possible futures; (a) *Communism*, as a combination of abundance and equality. (b) *Rentism* the idea that abundance combined with restricted access to goods (i.e. imposing artificial restrictions akin to copyright protections on music distribution). If we are to assume that technological progress cannot fully automate production, or, due to inherent limits, one cannot reach over-capacity in almost everything, a future with widespread equality would be called *Socialism*. The worst case scenario, termed *Exterminism*, projects a highly unequal future combined with scarcity, where the vast swaths of humanity being are condemned to irrelevance and base their reproduction by being servile to tiny fraction confined in reverse ghettos.

**3.2.3.2. Trajectories towards the limit case.** Automation of the level described in the previous paragraph seems to be far away – or at least not imminent. Technologies that originally seemed trivial and around the corner (e.g. self-driving cars, see Brooks, 2019) are now thought of as requiring years of further development. At this moment, AI seems to be automating jobs that were traditionally thought of as middle management; it also plays a role in intensifying labour. Examples include Uber's allocation algorithms and Amazon's hand gesture patents. The core of the issues lies with worker performance management increasingly delegated to machines (De Stefano, 2018), creating dystopia like conditions for the ones affected. It is hard to see how a full-automation (or almost full automation) society will not be impacted by the technological trajectory that led to it. It is also hard not to wonder when will this almost teleological point in history arrive. There seem to be wide disagreement among scholars and experts, but most agree we would have achieved full automation within 100 years (Müller & Bostrom, 2016; Walsh, 2018). If the current trends persist, the trajectory towards the AI limit case will be painful for most.

**3.2.3.3. What is missing?** The focus on management, surveillance and other labour disciplining technologies is not necessarily out of choice. We are ineffective at creating machines that act in unconstrained environments; our AI systems traditionally need copious amounts of data and tend to produce dubious results outside well defined conditions. They are also not very good at learning the invariances of this world, and fail to generalise outside their training distribution. This constraints practitioners into solving problems that can be attacked, which are almost always high level optimisation problems. Tasks like mending a broken car away from a production line, basic plumbing etc. are completely outside the capabilities of modern AI. The current crop of AI also fails to learn incrementally from data – something termed "catastrophic forgetting". One can only speculate what an AI for Good would imply, but my best guess is an always-on private advisor and personal guide. Efforts to develop similar technologies in the networking arena are late (see Bielenberg, Helm, Gentilucci, Stefanescu, & Zhang, 2012; Moglen, 2013), but have managed to gather significant traction – maybe there is room for an AI equivalent?

### 3.2.4. A generic perspective of AI – Arpan Kar

**3.2.4.1. Challenges.** The growing popularity of AI is changing the way firms are engaged in doing business across industries. The benefits and productivity improvement at the tactical, operational and strategic level are envisaged to be significant as the firms move towards digital transformation empowered with AI. The information assets residing in these firms, often referred to as big data, are systematically exploited and explored by AI to create this value (Grover & Kar, 2017). Similar value is also realised in government, public organisations and not for

profit organisations. However, there are major challenges remaining in this journey to explore and exploit the full potential of AI. In our limited perspectives, these can be segregated in terms of algorithm specific challenges, domain specific challenges and policy related challenges.

First, let me explain the algorithm specific challenges. Reviews of artificial intelligence algorithms (Chakraborty & Kar, 2016; Kar, 2016; Chakraborty & Kar, 2017) have demonstrated that most of these applications are focused on showcasing usage of relatively few approaches. Probably eighty percent or more of the published literature is covered by algorithms like neural networks, random forest and decision trees. To an extent some work has been published on algorithms like genetic algorithms and swarm intelligence. While around 2010, there was a new focus on developing new AI algorithms due to computational limitations of these age old algorithms, the focus has relatively died down with the growing popularity of deep and convoluted neural networks. This trend is predominantly driven by the growing access to high performance computing infrastructure in academic and industrial research units. Hence the focus on exploring newer algorithms for theory development has taken a back step as researchers are focusing on exploiting these established algorithms in the wake of improved computational infrastructure. So theory development in many niche algorithms has received less priority, as compared to what it should receive. This creates a gap in the long process of knowledge development in the domain, as current users continue to exploit well known algorithms with better computational platforms while exploration takes a backseat in this journey.

Next comes the domain specific challenges. The focus on exploiting AI inherently means that organisations have to systematically develop and maintain information assets, which requires a digital transformation within these organisations. However, in many of these organisations, when the initial planning is being done, there is a gap between digitalisation and digital transformation, due to the organisation's overall technology readiness. Therefore there is a challenge on identifying the right questions, what data needs to be captured to answer these questions, understanding approaches to extract, maintain and analyse these data and understanding the implications of this analysis. This journey essentially means that there has to be professionals who can understand both the functional elements of the organisation's processes as well as appreciate the technical elements of AI. This is often missing in existing Organisation as they gear towards taking a big leap to leverage AI, and thus sometimes, due to the lack of internal readiness, there is a productivity paradox that emerges due to lags in learning (Barton & Court, 2012). So AI usage needs to have greater adoption before it affects the industry productivity as a whole.

Further there are challenges of estimating trade-offs between differentiation versus commoditisation of AI. Like any information technology, the economic returns of AI is most high when it reaches a maturity of commoditisation. But AI systems are also expected to evolve themselves as they learn from the contextual and sticky knowledge within organisations, which mean that outcome of AI can never ever truly be commoditised. However current maturity of digital transformation journey across organisations, even within the same domain, does not facilitate too much of commoditisation.

Since it is a relatively new hype in terms of applications, although the technologies are old by quite a few decades, the actual readiness of the organisations producing or consuming the service enabled through AI, is often less understood. Process maturity and people maturity in such organisations become questionable. As a result, issues surrounding how these information assets are developed, maintained and exploited, becomes debatable, when adverse impacts are witnessed. Challenges of privacy preservation, security and process alignment becomes critical. Further people who will use these technologies need to undertake severe reskilling and deskilling.

On the policy side, similarly there is a lack of mature standards and public policy to address these challenges. What could be measures of intervention from government to control market concentration? The

organisations which are the market leaders in the segment have many customers onboarded, and as a result have the requisite data created in their platforms which is exploited by AI, to provide differentiating services. A small firm will be less effective purely due to lack of access to the data. However, this brings in a possibility of an AI divide where strong research units continue to grow and reap benefits while other smaller units fail to take their innovation to the next level. However, there are procedural challenges in the decentralisation of innovation in AI and policy to govern this.

**3.2.4.2. Opportunities.** The opportunities in AI are tremendous given that it is still at a rather nascent stage in terms of adoption in different industries. So it would be exciting for industries to explore different slices of AI impacts in different contexts. Such opportunities of theory development for researchers in AI could come from exploration in the following themes.

- Challenges and issues in managing AI in organisational and social setup.
- Impacts of AI on organisation design and associated issues
- Impacts on behaviour of individual stakeholders who are affected by AI and the cycle of how they affect the outcome of AI
- Requirements surrounding deskilling and reskilling human workforce in the wake of AI usage in Industry 4.0
- AI impacts from a systems methodology – drivers and actors of the ecosystem

Also it would be interesting to explore opportunities of how AI can be leveraged not only at the firm level but as an enabler in platforms and ecosystems. AI may help to connect multiple firms and help in automating and managing information flows across multiple organisations in such platforms. It would be good to explore opportunities for AI to be used in such platforms to impact platform productivity, firm productivity, and ecosystem productivity.

Further research in AI from a technical perspective would also have immense opportunities in the years to come.

- Areas could be related to computational algorithms for making sense out of unstructured and large volumes of data
- Exploration could be in non-deterministic polynomial-time hard problems even with structured data but high volumes
- It would also be interesting to explore the role of AI in conjunction with decision theories for management.

**3.2.4.3. Research agenda.** Building blocks for future research in AI can stem from integrating classic information systems research theories emerging from management theory, organisation theory, behavioural theory, computer science theories and systems theory (Barki, Rivard, & Talbot, 1993). This should be done by extending the exploration in computer science for contextual applications in organisations, markets and society. A particular area of interest would be the role of AI in networks consisting of different actors and types of linkages. Currently most of the research in AI is happening in computer science and information technology departments in universities. Such computational work is mostly getting presented in leading computer science conferences like Neural Information Processing Systems, International Conference on Machine Learning, Association for the Advancement of Artificial Intelligence and other such technical conferences. However the focus of such conferences is mostly addressing the computing block of information systems. Connecting such studies and progresses with management, organisation, behavioural and systems theories in information systems would allow exploration of multiple complex phenomenon of AI in this journey towards digital transformation. In particular, this may lead to very strong contribution for policy making and practice, based on such mixed research methods.

However there is also a strong need to relook at theory and relationships based on the emergence of AI. The nature of data available due to digital transformation is completely changing the traditional approaches of research. The presence of big data on platforms like shared economy, social media and internet of things, may require a different approach of theorising than from the traditional research methodologies (experiments, surveys and interviews). This journey would also require researchers design research methodologies from studies connecting both positivist and interpretive paradigms. For theory development, proxies of constructs may be derived from big data which is getting generated in different data sources. For example, relationships between such constructs may require both qualitative and econometric validation, and since the data has high veracity, new acceptable levels of statistical thresholds may require to be adopted for theory development. So there may be a perceived trade-off between rigour versus the generalisability and applicability of such findings if one were to evaluate such research purely from a reviewing perspective. However, given that the area is still nascent, it would be good overlook the minor rigour-relevance gaps and try to take forward some of the new insights in AI and their impacts on the ecosystem they serve.

### 3.2.5. Artificial Intelligence for digital marketing – Emmanuel Mogaji

The vast amount of data being generated, increased use of mobile device, cloud computing and internet has contributed to the significant development of Artificial Intelligence (AI). AI is making a double-edged impact – constituting a significant source of innovation yet threatening human jobs (Huang & Rust, 2018), this piece, however, focuses explicitly on discussing the opportunities of AI for digital marketing. Wirth (2018) noted that the industry seems hesitant and at the same time eager to embrace this new technology. This piece will explore some of the challenges for its adoption and research agenda for a better understanding.

**3.2.5.1. Opportunities.** Artificial intelligence offers opportunities to enhances campaign creation, planning, targeting, planning, and evaluation. Three key stakeholders are identified as the opportunities for AI in digital marketing are being explored. Firstly, the brands who need to understand their customers and communicate with them on a very personal and emotional level. Secondly, the Advertisers and Marketing agencies who are responsible for digital marketing strategies. They need AI to bridge the gap between the brands, the customers and data (Bell, 2019) and Thirdly, the customers who need to engage with the brands' marketing communications. They are the recipient of the information and the generator of the data which is being used to targeting. With this understanding, the opportunities for these stakeholders are presented, especially for digital marketers.

**Data:** A large amount of data generated by the consumers provides an insight into their behaviour Customer analytics makes up 48% of big data use in sales and marketing (Columbus, 2016) which highlight there are new sources of data about the customers. Advertisers have seized the opportunity to use this data to personalise and target advertisements (Boerman, Kruikemeier, & Zuiderveen Borgesius, 2017). Marketers have never had this form of data from the customers. No doubt these are big data collected over different touch points. However, AI offers the opportunity to process these data faster and effectively engage with everyone with messages that appeals to them. Segmentation and targeting become very easy through the data available.

**Content creation:** With AI being able to do what humans will typically do, there are opportunities for more innovative and relevant content creation. With consumers' demand for relevant content, advertisers can explore the prospects of AI to develop contents relevant to the customers because they now have a better understanding through the analysed data. Content here includes advertisement, social media posts and email campaigns. Information such as past purchases, interests and browsing behaviours can be used to create automated campaigns that can enhance the customers' purchase intention. AI can

identify the consumer's pattern about lifestyle choices including music, favourite celebrity and location to create unique content.

**Content sharing:** With the understanding of the customers and the creation of relevant content, AI can also go further to deliver these messages to the customers in a non-intrusive manner. With customers engagement and information collected such as location, demographics, devices, interaction with the site, AI can display offers and content that are more appropriate for each user type. Analysed information about the customers determine the best times and days of the week to send an email campaign or post on social media, the recommended frequency of the marketing messages and the title they are more likely to engage with. This content sharing opportunity builds on the power of Programmatic which allows automated bidding on advertising inventory in real time.

**3.2.5.2. Challenges.** Despite these opportunities, some challenges may hinder the adoption and implementation of AI for digital marketing.

**The availability of data:** There are challenges for collecting and using the data, especially considering The European Union General Data Protection Regulation (GDPR). Data is essential in understanding the customers, their journeys and developing the advertising campaigns. Personalised and automated content creation and sharing will not be possible if the data are not available. When customers are not willing to release relevant information, the AI algorithm is not receiving the needed resources to learn and develop the process. The AI algorithms need access to that data to give accurate recommendations. Even when the data is available, it should be AI ready, that is it is readily available for machine learning. Companies have been collecting information about their customers for many years, and it is essential that they start considering the information with regards to AI, making it structured enough for digital marketers to use.

**Resources:** Though AI is getting much attention as a fast-developing technology, the cost needed to it for digital marketing may be a limiting factor. Top IT companies and Start-ups champion most of these developments; it will not be surprised if AI is just limited to some of the biggest advertising agencies who have the financial capabilities. The financial implication of research and development that goes into it creating and maintaining AI does not make it readily available for everyone but no doubt it will become cheaper as times goes on. Besides, human resources needed to champion these projects might also pose a challenge. The level of knowledge about AI in digital marketing is not keeping pace with the developing in AI as it becomes increasingly sophisticated. The insufficient level of skill individuals may be a barrier to exploring the full capabilities of data-driven digital marketing.

**Trust in AI:** There are trust issues with AI (Siau & Wang, 2018). Advertising practitioners are feeling that the machine is not creative enough or it is going to take their jobs, Brand feelings they are losing grip over their narratives, allowing the machine to generate contents and not convinced the algorithms can deliver results. Consumers are feeling they are just being targeted. Knight (2017) suggested that there are dark secrets at the heart of AI because no one knows how the most advanced algorithms do what they do and that could be a problem. Presenting the state of AI in 2019, Vincent (2019) asked if computers are not explicitly taught (as they learn on their own), how do you know how they make decisions, he further argued that teaching computers to learn for themselves is a brilliant shortcut. Moreover, like all shortcuts, it involves cutting corners. This inherent fear about the prospect of AI highlight challenges for its adoption for marketing communications.

**3.2.5.3. Research agenda.** These opportunities and challenges open avenues for future research to understand how best to harness the prospects of AI within digital marketing.

**Conceptual and theoretical development:** AI has been applied in many different domains such as medicine, hospitality and travel. While the possibilities of using artificial intelligence (AI) to extract information about customers, generate advertisements that will appeal to them and



shared digitally has been presented, a holistic conceptual and theoretical understanding of these prospected is needed. The current hype around AI is creating a blurry picture calling for further research and clarification (Wirth, 2018). A systematic review of AI-related applications in digital marketing, definitions and terms with empirical insight is needed, especially within the context of the stakeholders – the advertisers, brands and consumers. Exploring how firms should integrate AI, either as human replacements or integration (Huang & Rust, 2018).

*AI integration with OBA and MLBA:* Following on the conceptual and theoretical development of AI in digital marketing, the integration with online behavioural advertising (OBA) and mobile location-based advertising (MLBA). Currently, behavioural targeting mostly occurs when using computers or smartphones (Gutierrez, O'Leary, Rana, Dwivedi, & Calle, 2019), scholars argue that it offers personalised and targeted advertisement, offering a precise way of targeting customers (Kumar & Shaphali, 2016) and contributing to the growth in Online advertising revenues (Chen & Stallaert, 2014). Likewise, MLBA offers consumers benefits such as personalised communications that are tailored to the mobile user's real-time geographic location (Krishen, Raschke, Close, & Kachroo, 2017). These two concepts are emerging marketing strategies, and it involves collecting data either online or off-line and using it to develop advertising campaigns. With AI offering data collection and processing at a faster rate, a better understand and effort towards triangulating these online and offline data to have a better understanding of consumers' is essential. Providing practical implications for marketing researchers and practitioners.

*Ethics:* The ethical consideration between the personalisation of advertising and consumers' privacy represents the personalisation-privacy paradox (PPP) (Gutierrez et al., 2019) often explored through the privacy calculus theory (PCT) (Xu et al., 2019). Collecting, using, and sharing personal data for marketing purposed has always raised consumer privacy concerns (Boerman et al., 2017), this concern is expected to grow as machines are being deployed to extract and process these data. A better understanding of how stakeholders deal with this concern is important. Consumers have misconceptions about extracting data for marketing purposes as they have little knowledge about it (Smit, Van Noort, & Voorveld, 2014) and yet advertisers and brands keep extracting these data. This is what Boerman et al. (2017) described as 'information asymmetry' where companies know much about consumers, yet consumers know little about what happens to their data. Considering machine extracting these data without human intervention, a theoretical understanding of its implication is worth considering.

*Content creation:* It will be necessary to intensify the exploration of the content creation capabilities of AI for digital marketing. This is an agenda marketing practitioner will find relevant, this follows the conclusion by Mogaji, Olaleye, and Ukpabi (2018) that personal data and information legitimately collected online by companies can be used to design and personalise advertisements that appeals to consumers emotions and shared online. This process differs from the highly personalised and rational data such as age, gender, and location which Aguirre et al. (2015) found to have a reduced click-through rate but things that emotionally appeals to individuals like their choice of colours, images being used and background music. Harnessing data and expertise offered by AI to develop the marketing strategy offered a and enhance decision-making process, as Boerman et al. (2017) advised that advertisers should consider the level of personalisation as advertisements perceived to be too personal can seem intrusive

*Attitude towards AI developed campaigns:* There have been previous studies that uncovered some favourable and unfavourable consumer responses to ad personalisation, but the moderating factors that strengthen or weaken these effects are still mostly missing (Bleier & Eisenbeiss, 2015). Building on the previous study that explored consumer avoidance of personalised advertising (Baek & Morimoto, 2012), future research should endeavour to empirically explore factors that can influence the acceptance and avoidance of AI in digital marketing. Research should consider how well AI is targeting the customers with

relevant advertisements as personalised advertisements increase intention to purchase when advertisement fits customers need (Van Doorn & Hoekstra, 2013). AI can extract the data and deliver the advertisement but how well are consumers engaging with it? What are they engaging with and what are they finding appealing. This information can help shape future development and ergonomic design of AI systems.

*Stakeholders attitude towards AI for Digital marketing:* Advertisers, consumers and brands attitude towards the innovation is also worth researching. As (Huang & Rust, 2018) noted, AI is taking over human jobs, come creative tasks are being threatened. How are professionally able to deal with these dynamics? How will marketers and advertising practitioners integrate AI into their job? Privacy concerns of consumers and trust in AI, transparency about the reason why companies and advertising agencies are collecting the data could be some deterrent to this initiative as Jai, Burns, and King (2013) found that when consumers know that their information, collected on websites are shared with third-party companies, there is lower repurchase intentions, increased perceived risk and unfairness. Brands are aware of the financial implication and the benefit, how eager are they to explore the prospects of AI?

*Metrics and evaluation:* The relationship between advertisement and intention to purchase should be revisited in the context of AI in digital marketing. Consumer awareness of personalised targeting through extracted data alter consumers' responses to online behavioural advertisements (Aguirre et al., 2015). Also, Humans beings may be difficult to monitor and observed (Mogaji et al., 2018). Their browsing history may not be a true reflection of their personality and what appeals to their emotions. Therefore, the metrics and form of evaluating the effectiveness of the AI developed campaign should be further explored. There is a need to develop and test the practically feasible of AI impact, its contribution towards the industry's growth, if it has increased sales for brands and if it has enhanced consumers choice making process.

*Conclusion:* While acknowledging the role and advancement of AI in everyday life, this piece has focused explicitly on the role of AI in digital marketing. The prospects, challenges and research agenda has been explored. AI offers enormous opportunities for key stakeholders. AI helps marketing agencies gets a better understanding of the data, to meet their goals and help brands connects emotionally to their customers. AI open opportunity for interdisciplinary collaboration involving AI developers and creative individuals, enhancing the power of AI to develop appealing advertising campaigns. Researchers following the research agenda provide theoretical insight and managerial implications relevant for AI developers, marketers and brand managers.

### 3.2.6. Artificial Intelligence for sales – Kenneth Le Meunier-Fitzhugh and Leslie Caroline Le Meunier-FitzHugh

AI is changing the business landscape, and its effects are no less in Sales than in any other business function. Sales is where the business 'meets' the customer, whether it is in retail situations (business-to-consumer, B2C) or in business-to-business sales (B2B). Sales can take place face-to-face, through retail outlets, over the Internet or other communication media. Retail B2C shopping via platforms such as Amazon, is guided and influenced by various AI algorithms that have completely changed the retail selling experiences and this trend is set to continue for the foreseeable future. For example, recommendation algorithms present 'suitable' offers to on-line customers to consider, rather than waiting for the customer to make their selection. Additionally, a location algorithm will allow the customer to be presented with the location of outlets linked to their sales selections (Antonio, 2018). However, the impact of AI on sales generally goes much further than this. B2B sales are often more complex and have a greater monetary value per exchange than retail sales, although the latter are more numerous. The challenge for B2B sales is to understand how AI is influencing sales exchanges. The day has already come where salespeople may be prompted by AI in real time via their tablet or phone during



sales negotiations, in both B2B and B2C interactions. Further, some B2B sales functions are already being automated e.g. customer relationship software being used to identify sales readiness and telesales calls being initiated and guided by automated systems. It could be just a short step to AI providing integrative sales experiences on-line and through tele-services that could remove the need for salespeople altogether. The following section will explore some of the challenges and opportunities being offered by AI around the sales function, and then presents future possible research questions.

**3.2.6.1. Emerging challenges. AI is a job killer:** Automation and automated services are replacing people in many industrial and service sectors to reduce costs and improve efficiencies. It is forecast that over 375 million jobs globally may be 'lost' to AI by 2030 (<https://www.scientificworldinfo.com>). The effectiveness of automation in sales has already been proven in telesales and forecasting situations, which have resulted in multiple sales-related job losses. It may be that as AI develops the mechanistic nature of the algorithms may reduce the need for competitive differentiation and creativity in sales exchanges. Additionally, as competitors use similar information derived from big data analytics to base their new product development (NPD) decisions on, or to develop 'customised' offers, it may become harder for customers to differentiate between offers and for salespeople to justify their existence. The proposition is that increased use of AI could reduce investment in sales training and development, increase monopoly/oligopoly situations and reduce the sales work available in that industry.

**Loss of privacy:** The increased use of automated systems and AI opens up customers to privacy risks. While CRM systems allow companies to compare their customers' buying behaviours, the customers are at risk of finding this information in the public domain or being 'sold' to other interested parties. Salespeople may find that their customers are being 'poached' by other salespeople who have access to automatically-generated leads. Alternatively, salespeople may find that their previously personalised data that forms the bases of their customer relationships is now readily available to other parts of the company, reducing their effectiveness.

**Changing the nature of salespeople's interactions with customers:** The danger of the increasing use of AI allows B2B customer to self-select their sales deals, including pricing, discounts and special features, removing the need for human interaction. However, increasing our reliance on AI selling systems may result in Trust in the selling organisation being damaged. The increased use of AI removes the human touch and emotional connectivity that customers have developed with their sales representative. This loss of authenticity in the sales process may mean that buyers are unsure who to trust (Hurley, 2017). A further danger is that basing sales forecasting on past behaviours and on-line activities may perpetuate a bias, e.g. represent a past or random concern that is no longer relevant to the customer's current buying needs. While new types of customer insights may be created through AI, some of these may be too detailed or on the wrong track to be profitable. Finally, there is a danger of losing control of the sales process as integrated, automated systems prioritises emails, tracks new contacts and creates meetings or agendas of which the salesperson is unaware. This may lead to the salesperson no longer feeling in control of their activities and worrying about missing new opportunities that are not highlighted through the automated system (Loring, 2018), but which may have been identified by the salesperson in their interactions.

**Lack of understanding of AI algorithms:** It is sensible to pose the question, exactly what do sales algorithms do? Algorithms are meant to provide guidance and information to sales action by moving the customer along a prescribed route (Knight, 2017). They collect data on what the customer is looking at or is doing, and extrapolates the information that salespeople might require, which may be very helpful, or extremely annoying. The effects of algorithms still need to be researched further. If AI is driving market research, can the results be

trusted? With the increase use of bots and auto-response algorithms, how can on-line market research be verified? The salesperson may be presented with inferior or bias information on which to base their sales negotiations.

**3.2.6.2. Opportunities. Managing performance of salespeople:** AI provides Sales Managers with dynamic assessments of performance via AI driven dashboards. They can be used to identify upselling and cross-selling opportunities to the company's customers. AI may be also be used as a personal assistant that is able to schedule meetings with selected partners, releasing the salesperson's time for actually attending the meetings. An AI assistant can pull through prospective and existing customer data from internal files against key criteria or specified names (Loring, 2018).

**AI predictive abilities in sales forecasting and customer management:** Salespeople love up-to-date data continuously streamed to them wherever they are based. Customers like to be personally communicated with and AI can help salespeople to develop their relationships by providing this personalised information, saving time and preventing mis-directed sales efforts. Combining new algorithms with existing CRM platforms should allow for the analyses and prediction of selling opportunities, or the salesperson to identify changes in customer status (Antonio, 2018). The ability to leverage big data to focus the sales professional on their target customers, should enable building more authentic relationships. Additionally, scenarios and coaching may be provided by advanced behavioural analytics to produce suggestions of how to handle blockages in the sales process, and how to benchmark themselves against top performers (Hurley, 2017). An AI system can automatically update the CRM system through monitoring incoming and outgoing data. The predictive capabilities of AI can also now be used to gauge the customer's possible lifetime value, allowing salespeople to invest in these growth areas and to offer incentives and interactions geared to the customer's individual needs (Loring, 2018). Behavioural analytics and customer profiling may be used to provide salespeople with the ability to personalise interactions to meet the needs of their different customers. Some systems will also have the ability to analyse conversations to identify approaches that increases effectiveness in handling interactions and greater sales successes (Hurley, 2017).

**The effects of big data analysis on prospecting for new leads and customer retention:** Managing big data is a key area that AI can help to streamline. The prospecting process for new customers is time consuming and frequently leads to disappointment. The AI algorithm can provide an interaction history based on contacts and social media activities (Antonio, 2018) and AI led lead-scoring may be used to identify who is ready to buy, and who in the pipeline is reading to move from prospector to customer. Using AI identified leads enables salespeople to concentrate on a significant number of potentially beneficial sales leads that, once qualified, can be guided through the buying journey towards purchase (Loring, 2018). Therefore, by employing AI systems to identify patterns in customer behaviours it is possible to pinpoint customers in the sales pipeline who are at sales readiness, as well as identifying buying trends in existing markets and the possible emergence of new markets. Unstructured data can be analysed, manipulated and presented in a structured way, so it may be used in the sales process and help simplify sales conversations through the provision of key facts about the customer's interests (<https://www.scientificworldinfo.com>). Salespeople may also use AI systems to improve customer retention and optimisation (Loring, 2018). Buyers are kept loyal with customised incentive programmes. AI-assisted customer care programmes will help to make sure that the salesperson is aligned with their customer's needs and current interests. Real time purchases in-store and on-line may be increased through point of sales information being provided through conversational interfaces and virtual assistants providing the right information at critical points in the sales process (<https://www.scientificworldinfo.com>).

*The effect of AI systems on Salespeople's cognitive and professional development:* The effects of AI is challenging salespeople to develop new knowledge, skills, including management roles (Loten, 2017). The use of AI algorithms is contributing to productivity and provides sales process enhancement through elimination of non-productive activities and through the removal of mundane jobs. AI can also speed up the sales process by identifying changes in buying patterns and taking over some of the more repetitive administrative roles (Loring, 2018). Salespeople can then concentrate on building relationships with the customers that are identified as having the greatest potential growth and lifetime value by their AI assistant.

### 3.2.6.3. Research questions

- How far should AI be encouraged to take the sales process – human buyer interacting with AI sales assistants, or AI buyers purchasing from AI sales assistants?
- How can big data analysis of social media and other on-line sources be used to allow the salesperson to develop greater adaptability to their customer's needs?
- Will AI free the salesperson to develop higher level, soft evaluative skills and emotional intelligence that are required to handle high level, personalised sales interactions, or consign salespeople to mundane, data driven interactions?
- Will AI standardise sales performance across industries or allow for the development of more individualised sales roles?
- How can privacy and sales intelligence be safe-guarded in an AI driven environment?
- What are the effects of algorithms on customers' behaviour?

### 3.2.7. Complementary assets and affordable-tech as pathways for AI in the developing world: case of India – P. Vigneswara Ilavarasan

This section argues that pathways for AI are complementary assets and affordable-tech in a developing country. The development, implementation and adoption of AI are dependent on the first two pathways. The future research agenda should explore possibilities of AI based affordable technologies. This argument is discussed in the context of the developing world countries, using a case study of India.

Studies on the adoption, use and impact of information technology (IT) in businesses or organisations are not new (for instance Bili & Raymond, 1993; Brown, 2015; Liu, Ke, Wei, & Hua, 2013). The value extraction from IT investments is enabled by the presence of complementary assets (Hughes & Morton, 2006). The complementary assets can be discussed in three major areas ((Laudon & Laudon, 2017): organisational (culture, structure, process and people resources); managerial (top management support, incentives to use, and training); and social (national IT and telecom infrastructure, education, supportive regulatory environment and legislators, and technology business ecosystem).

In a wide range of areas AI can be deployed (please see for a recent review, Duan et al., 2019). For instance, Ransbotham, Kiron, Gerbert, and Reeves (2017) highlights that AI reduced the production rate in Air Bus. A quick review (Dutton, 2018) showed that high income countries have either invested money for AI related research or have established institutions to explore the potential benefits. The review also indicates that high income countries are leading the AI bandwagon. Except, India, Kenya and Tunisia, none of the developing world countries are part of this discourse. The presence and absence of complementary assets are likely to influence the trajectories of AI in the developing world. The lessons learnt from India or Kenya experiences are likely to benefit similar low income countries in future. For instance, Pakistan, Vietnam, and others are imitating the success story of Indian software services exports story (Steinmueller, 2001; Yim et al., 2016).

For discussion purposes, we will look at three areas of complementary assets – national IT and telecom infrastructure; nature of enterprises; and regulatory and legislative environment. The national

IT and telecom infrastructure of the world is captured by the three indices: E-Government Development Index (United Nations, 2018), Information and Communication Technologies (ICT) Development Index (ITU, 2017), and Networked Readiness Index (Baller, Dutta, & Lanvin, 2016). The ranks scored by India are 96, 143 and 134 respectively. These three indices include parameters that reflect the quality of complementary assets. Some of the parameters are Internet users, households and individual subscribers of telephone and broadband (fixed and mobile) per 100 inhabitants, firm level IT adoption, online public services availability and quality of relevant people resources available. As India is doing poorly in the parameters, the overall ranking is low inferring the nature of complementary assets available for the AI domain.

The second area of complementary assets is the nature of enterprises in India. An official estimate (Data Gov, 2018) shows that in 2011 there were 57.6 million enterprises in India. Out of the total, 84.5% are own account enterprises. Though the enterprises are employing a larger volume of people, the nature of enterprises indicates poor technological capabilities and small size. An analysis of recent data on employment published by the government of India (Ilavarasan, 2018) showed that nearly two-thirds of the workforce is employed in small size enterprises. More than half of them are working in businesses with less than six employees. The poor technological capability of these enterprises can be inferred by the fact that two-third of them do not use electricity.

More than two thirds of enterprises are owned. The enterprise owners are likely to be less educated. If AI technology is made available, the absorptive capacity of these owners ability to understand the potential is limited (Roberts, Galluch, Dinger, & Grover, 2012). In India, less than one quarter of firms are using AI in their business processes, and start-up ecosystem in the AI domain is miniscule (Niti Aayog, 2018). Not surprisingly, even in the USA, only 17% of the 1500 senior executives surveyed had an understanding of AI and its applications in their businesses (Bradbury, 2018). If AI deployment is possible only in enterprises that are larger with a threshold amount of technological capabilities, the scale of adoption in future will be low in India.

The regulatory and legislative environment in India is positive. The digitisation led policy initiatives collectively called as Digital India pursued by the Government of India are proactive towards AI (Niti Aayog, 2018). The Niti Aayog, the policy recommendation body of Govt. of India is keen on using AI for the national development. The areas designated for change are healthcare, agriculture, education, smart cities and infrastructure and smart mobility and transportation. Out of the listed areas, it is estimated that US\$ 14 billion investment is committed to creating 100 smart cities in India (Pratap, 2015). The government recognises the inadequacies in its machinery. The policy document seeks the private players to participate in the AI development and deployment including the readying the people resources. It is also open to working with others to develop AI based solutions. For instance, Wadhvani AI (<https://eng.wadhwaniai.org/>), a non-profit research institution connects AI experts from universities, grassroots non-governmental organisations and government organisations in developing AI based solutions for social good. At present, the work domains are maternal and child health, tuberculosis and cotton farming in India.

However, legislators are likely to prefer AI developments that do not replace people. Frey and Osborne (2017) postulated that AI based automation technologies are likely to displace jobs, even those include cognitive tasks, in the world. Based on their work, there are predictions that 52% (McKinsey Global Institute, 2017) to 77% (Ilavarasan, 2018) of the jobs in India will be automated shortly. Given that the average family, size is four, the impact of automation shall have serious implications. As women and other disadvantaged groups are predominantly doing low skilled jobs, they are also likely to be replaced by AI based technologies (Ilavarasan, 2018). The consequences of the labour displacement might include severe social disorder. No wonder, Mr Gadkari, Union minister has publicly announced that driverless cars shall not be permitted in India<sup>2</sup> signalling no support for AI in similar

lines.

In light of the above, we hypothesise that complementary assets are weak or absent for AI in the low income countries. This is the major challenge for the widespread AI adoption and usage in the low income countries. This challenge can be addressed if the AI based solutions are affordable and cater to low technology enterprises or users. There are some positive demonstrations from both small and larger firms. For instance, Get My Parking, a mobile application based solution is being used by low educated parking attendants or contractors in parking spaces in New Delhi.<sup>3</sup> This application requires low capital and short learning curves. The use of applications like UBER or Google assistant by the drivers in regional languages in their low cost smart phones indicates the adoption of AI based technologies. Get My Parking is a technology start-up whereas Google is a larger technology firm. Small firms are handicapped from lack of access to a larger volume of data which is essential for refining AI solutions, but likely to target markets not catered by the biggies. However, we do not know whether AI ecosystem is dominated by the start-ups or resemble oligopolistic arrangements.

In the light of the above, future research can focus on the following questions:

- What are country specific factors that drive the development and deployment of AI? Also, why India and Kenya are few among the developing countries are exploring the potential of AI than others?
- How do different complementary assets result in divergent AI adoption and deployment levels?
- Do the small firm's dominant industrial structure hinder the AI development and growth?
- Whether the likely users, both from industry and government, possess adequate knowledge about AI and its applications? How does this awareness affect the level of adoption and support for AI ecosystem?
- Whether affordability is the prime factor for the adoption of AI in the developing countries?
- Whether AI and impact on employment linkages is a repeat of old debate on computerisation and labour displacement?
- How do national governments balance the good social potential of AI vis-vis the potential negative consequences for the future of work?
- How can private technology firms collaborate with the national governments in the developing world in developing the required skills and deploying AI based solutions?
- How do start-ups compete with the larger firms in the AI market? Do they focus on different application domains? How do we explain or understand the differences?

### 3.3. Arts, humanities & law perspective

#### 3.3.1. People-centred perspectives on Artificial Intelligence – Jak Spencer

Whilst AI has a multitude of different technological, political and legislative challenges and opportunities, ultimately it has the most impact on people and their everyday lives. People-centred design is a form of innovation that starts with empathy for people and ends with iterative solutions to solve real people's needs. Inclusive Design uses this methodology to create new innovations that meet the needs of the widest number of people as possible, no matter their age, ability, social, cultural or economical background (Clarkson, Coleman, Keates, & Lebbon, 2003). In recent years, many projects at the Helen Hamlyn Centre for Design, based at the Royal College of Art in London, have

focused on the impact and implications of AI in peoples' lives. The challenges, opportunities and future research agenda are outlined below.

**3.3.1.1. Challenges.** One of the main challenges regarding the adoption and implementation of AI is the current connotations and perceptions of the subject. To most people, AI is a mysterious concept that is not only hard to define, but also difficult to understand how it manifests itself in their everyday lives. Whilst obviously there are numerous positive uses of AI, many people associate it with negative press and media campaigns depicting AI as the cause of everything from mass unemployment to data breaches, removal of freedom and even full-out global warfare. The ill-defined concept and poor media coverage has given AI a negative brand image, and the jury is still out as to whether the good outweighs the bad.

Once you delve into more detail, further issues arise. The prevalence of biases are now well documented in many of the forms of AI we interact with, from racist financial algorithms to sexist chatbots (O'Neil, 2016). Part of the problem is the controversial role of playing the 'creator'. Artificial Intelligence that mimics human relationships have been created by people with a set of preconceived judgments, moralities, ethics and biases. As O'Neil states "algorithms are just opinions embedded in code" (O'Neil, 2016). Even as we move towards true intelligent robots that can build themselves, they still use human culture as a source for understanding relationships; stereotypes, discrimination, prejudices and all. In one recent example an experimental conversational agent that learnt from Twitter conversations took less than 24 h before it started tweeting hateful, racist, sexist and homophobic phrases (Vincent, 2016). One recent study by researchers at the Helen Hamlyn Centre for Design (Spencer, Poggi, & Gheerawo, 2018) found that the vast majority of digital assistants are portrayed as young, Caucasian, women that enhance the negative perception of the stereotyped role of women in secretarial roles. This is further enhanced by the language used by virtual assistants in general conversation and when responding to certain types of harassment (in one study, none of the major virtual assistants responded in a negative way to being called a slut (Fessler, 2017)).

Of course, many of these issues can be resolved by first questioning why we are looking to develop AI for a particular problem, and what will be the implications for people. At the moment, Artificial Intelligence is predominantly concerned with activities of productivity, efficiency and advancement of business objectives and ultimately increasing growth in the financial sense. In the very near future, we'll be able to have a cup of sugar delivered to our home by drones, work in a completely virtual office with efficient and productive artificial colleagues, or have the latest fashion trend waiting in our size in the wardrobe as soon as we get home. But what if this 'efficient living' isn't entirely a good thing? Many studies have shown that social interaction is fundamental to maintaining good mental health, whilst discovery, taking notice of things around us and serendipity are also important. In Japan, the term hikikomori refers to people who shut themselves off from society, often never leaving the house, relying on one or two close family members, or deliveries from online stores to sustain their lives. In 2016, Japan had 540,000 people aged between 15 and 39 who had not left their homes in last year (Ma, 2018). People can now live their entire existence without leaving the confines of their own home – and this must be having significant effects on our mental health.

A final note of caution amongst the challenges of new AI is the transition phase between things working with AI capability, and things working with reduced 'smart' capabilities – a phase termed 'augmented intelligence' we are already seeing the consequences of. At best, this transition phase can be confusing and frustrating, at worst discriminating and exclusive. One challenge remains how 'natural' the interaction with AI can be. Whilst in the future, this may be seamless, during this transition phase interactions are still on the terms of what the machine can understand. A recent research participant suffering

<sup>2</sup> <https://auto.ndtv.com/news/decided-not-to-allow-driverless-cars-in-india-nitin-gadkari-1964489>.

<sup>3</sup> <https://yourstory.com/2018/01/power-iot-get-parking-wants-disrupt-parking-industry-make-driving-easier>.

from early on-set dementia had been bought a smart speaker digital assistant by his daughter to help him with daily activities and reminders. Ultimately it went unused because the participant wanted to 'schedule an appointment in his diary' whilst the digital assistant only understood 'booking an event in the calendar'. In this instance, whilst the technology was advertised as an almost human-like assistant, in reality, users have to learn a new form of interaction, moving from the visual language we use on screens to verbal interactions through speakers and microphones.

**3.3.1.2. Opportunities.** Of course, the fact that so many challenges remain can also be interpreted positively with the amount of opportunity there is to improve our interaction and relationship with AI. One obvious opportunity that is already being explored, is the use of AI to enhance 'softer' goals rather than the persistent drive for economic productivity or financial efficiency. Indeed, this is one area where people-centred, qualitative research can really enhance further technological advancements by identifying genuine needs of people to deliver innovations with relevancy that solve real-world problems. With virtual assistants for example, conducting people-centred design research with a small number of diverse people resulted in the creation of an assistant that moved away from the generic stereotypical young, Caucasian, secretarial woman to the development of an older, wiser personality such as David Attenborough or Judy Dench (Spencer et al., 2018). As our interactions with machines start to become more and more human-like, the opportunity lies in the design of these new personalities and the creation of new types of relationship. Should these interactions be on a peer-level, subservient-level or superior-level? Should we treat new technology in the same way we treat our friends, bosses, parents or pets? These are questions that surely have to be answered on a case-by-case basis, and design can help to do this.

The potential of AI to help solve some of the world's most pressing social challenges is also one that cannot be overlooked. From our ageing population to the loneliness epidemic, from quicker and more accurate healthcare diagnostics to poverty alleviation, AI is helping to shape major global social challenges. Again, to be impactful here is to combine big data with the deep data of design. Algorithms and intelligence can be designed by engineers and computer scientists, but involving designers who are trained in understanding the needs, frustrations, behaviours and attitudes of real people can provide more innovative and ultimately impactful solutions.

A further opportunity lies in the 're-branding' of AI to something people can appreciate and even rally behind. Often the negative stories relating to unequal, biased or discriminatory algorithms are due to the opaque nature of both the definition and the underlying functionality of AI being misunderstood. Involving designers in the process and involving people helps to create more open, fair and even democratised AI that serves the people.

**3.3.1.3. Research agenda.** One of the most pressing challenges in the development of future AI is ensuring that it is not developed in a silo, without input from other disciplines, and importantly, real people. People-centred design has a history of acting as the 'cement between the bricks' of more 'scientific' disciplines and can generate impactful innovations in combination with AI. However, more understanding is needed on how the two disciplines can come together and combine in the most fruitful ways. This needs to be not only on a process level – combining two distinctly different disciplines, but also at an organisational and educational level – ensuring that teams dedicated to solving serious challenges have a mixture of disciplines and can develop and disseminate the mixed methodologies they use. The recent acquisition of Datascope by people-centred design firm IDEO, as well as the number of design and data labs being set up by the 'big 4' consultancies go some way to realising this, but future dissemination of successful and unsuccessful methodologies is important.

Another important future direction is ensuring that people are at the centre of any AI developments. In the practical sense, this means moving beyond AI for efficiency, towards creating more fair, just and equitable uses that not only improve people's lives, but also go on to enhance them in the form of creating pleasurable experiences – joy, connection, play and laughter. Transparency can help with this by reducing the mystique and opaqueness of AI to the general public. Acceptance of failure is also important – there will be mistakes along the way, but we need to be accepting of this and learn from them.

There is still a huge amount of research that needs to be completed on the ethical challenges of introducing new technologies into our lives. What level of responsibility do we give these new machines and what level of blame do they get when something goes wrong? Is it fair to ask machines questions we would struggle to answer as humans? One question that is often asked in the driverless car debate is how an autonomous vehicle would decide on whom to hit – an 80-year-old or a 3-year-old in the worst-case scenario (Awad et al., 2018). But is this something we could answer ourselves? It also opens up other areas of research enquiry, such as brand challenges of AI. A project at the Helen Hamlyn Centre recently reached out to a major global tech company to explore the AI possibility of helping to reduce suicides along a river-front, but discussions stalled because of the brand implications of being involved in such a sensitive topic and how it might look, despite the success or failure of the technology.

### *3.3.2. Taste, fear and cultural proximity in the demand for AI goods and services – Annie Tubadji*

All industrial revolutions generated not only economic but also important social challenges and opportunities, and yet there is currently no scientific economic recognition of the importance of social changes that may emerge due to the fourth industrial revolution. AI-generated goods and services might be objectively more efficient and less costly than human-made ones, yet we know from the hedonic valuation and behavioural economics literatures that human taste is not based entirely only on the objective characteristics of a product or service. People generally fear what is unknown and new because it brings them feelings of uncertainty. For the same reason they tend to prefer what is culturally closer to themselves. Taste, fear and cultural proximity seem to cause significant biases in consumer behaviour and this will inevitably affect the demand for AI goods and services and yet this issue remains under-researched. Meanwhile, whether AI-induced social changes are palatable to individuals and society will determine whether there will be demand and therefore whether a market will exist for AI-goods and services.

The first industrial revolution brought the steam engine and it helped travel between continents, trade intensified the growth of cities and slavery was abolished, and laws and institutions evolved. The second industrial revolution brought mass production and the car, which increased urban sprawl, cities grew into megacities, wages increased because factories were more efficient and work times got shorter and all this gradually allowed for the growth of the middle class. The third industrial revolution with the invention of the computer brought increased automation and productivity but also increased pollution as a social aftermath (Kling, 1991; Langton, 1984). The literature on the fourth industrial revolution however has remained narrowly focused on the question of skills and whether human and AI are substitutes or complements on the labour market (Acemoglu, Autor, Dorn, Hanson, & Price, 2014; Autor, 2013; Brynjolfsson & Hitt, 2000; Frey & Osborne, 2017; Genz, Lehmer, & Janser, 2018; Katz & Margo, 2013). The literature is yet to engage with the significance of social change that will be involved in switching between human-made and AI-made goods and services. AI is intensively used for learning about human demand patterns to serve as a technology that enhances the efficiency of marketing Kwong, Jiang, and Luo (2016) and retailing (Weber & Schütte, 2019). Very few studies have paid much attention to the specificities of the demand for AI goods and services, but two



examples do exist. The first is a quasi-insight on the demand for goods and services: Morikawa (2017) examines individual predictions about the type of industry that is likely (and thus inferred by the authors to be desirable or more acceptable for the interviewee) to experience substitution of human with AI labour. The other study focuses on the demand for AI services in a very specialised field: Tubadji, Webber, & Denny (2019) look at the demand for robo-advisory services in the banking sector. Apart from these two studies, the topic is largely ignored in the economic literature in favour of supply side research on how producers will be interested in the trade-off between human and AI workers. Yet, the challenges and opportunities of AI-induced social change may have complex implications for the diffusion of AI goods and services and therefore deserve careful examination.

We classify the challenges and opportunities for AI-goods and services in relation to social change and consumer behaviour into three categories: (i) taste, (ii) fear and (iii) cultural proximity. These are overviewed below as follows.

**3.3.2.1. Challenges.** Taste is a challenge that hides a double-edged sword. On one side, people adapt relatively slowly and definitely slower than artificial intelligence. The first appearance of the locomotive on film scared people and it took time until cinema established itself as a successful industry. AI might be more efficient and reliable, but using these types of goods and services may face social resistance at least for some considerable time after these services will be possible to offer on the market (Patsiotis, Hughes, & Webber, 2012). For a market to at all exist, both demand and supply are necessary. On the other side, taste is an unfaithful friend: over-enthusiastic but unjustified embracing of certain products and fashionable tastes creates the famous price bubbles such as the tulip mania. First, just like in the tulip mania bubble, bubbles lead first to massive precarious economic conditions. In the Netherlands, many individuals started to risk their entire economic fortune on obtaining a single tulip bulb. The situation aggravated so much that public policy interventions were due to prevent a national disaster (Garber, 2000). Second, taste-driven price bubbles tend to 'burst', i.e. price booms are famously followed by deep busts, which are proportional to the size of the boom (Breuninger & Berg, 2001; Emmett, 2000). Thus, if AI is accepted with enthusiastic fashion, it might explosively spread before the realistic aftermaths from its use get properly perceived and reflected by the demand for AI goods and services.

**Fear** is known in behavioural economics and prospect theory to be a stronger driver for human choice than pleasure (Kahneman & Tversky, 1979). While AI might be able to decrease our working hours and offer more efficient services, fear about competition for jobs between humans and AI and fear about the unknown impact of AI may lead us to underestimate the gains from AI and overestimate the threats. Thus, we may remove AI from our options for choice much earlier than we should (Shackle, 1949). An interviewee's severity of fear (as opposed to objective knowledge) about AI has not been seriously examined by either Morikawa (2017) or Frey and Osborne (2017). Meanwhile, Tubadji et al. (2019) find that increased experience in using technology is actually associated with an individual's disenchantment with the use of robo-advisory services in the banking sector. This finding suggests that even when fear is not initially present, a backlash may originate afterwards from the initial over-optimism about AI goods and services.

**Cultural proximity** might be the ultimate challenge for the embracing of AI goods and services. The goods and services of the first, second and third industrial revolutions were tools that enhanced what remained a predominantly human production. Steamboats moved people faster across the sea; the automobile made the suburbs closer in time to the city centre for the urban worker; computers connected humans. The negative social effect of Facebook and other known electronic media is based on the fact that they dehumanise social experience and people become less social. People are parochially thinking social animals that tend to exhibit preference for homogeneity, i.e. to show strong

preferences for things familiar and similar to one's own identity (McPherson, Smith-Lovin, & Cook, 2001). This latter tendency is often labelled as cultural proximity and is widely researched in the economic literature from Adam Smith to modern regional economic spatial analysis (Torre, 2008). Cultural proximity is known to be a key determinant for economic flows for people, goods, financial investments in gravity models – people prefer to live with, trade with and invest in what is closer to their own identity (i.e. the home bias effect) (Tadesse & White, 2010). AI lack human identity, so they are by definition at one degree distance from any human. How will this cultural distance between humans and AI interact with people's demand for AI goods and services? Are AI-products real substitutes for human product according to consumers' perceptions? Which products and services will be affected most from the lack of human proximity between the labour employed in these goods and service and the consumer?

**3.3.2.2. Opportunities.** Learning about the market *taste* for AI-produced goods and services can firstly improve firms and employers use of AI. While their attention is currently focused on the pros from efficiency gains, they might be overlooking the market reaction to the integration of AI in their production process. Learning about tastes informs the market about AI-generated products and services, which can help determine the quantity of AI-goods and services to invested in. This can prevent producers from being trapped with over-investment in a type of technology whose product might not be readily accepted by consumers. Taste-studies can also signal to the producer if the current high demand is a temporary bubble.

Learning about the *fear* factor in AI-related social opinions and policy-making tendencies can help us make evidence-based AI-related decisions. It will save us from being swayed by bounded rationality in our economic, political and policy-making decisions that relate to this novel technological device.

Learning about the importance of *cultural proximity* in the context of AI-human cultural distance can help to quantify the cultural gravity effect that bounds our consumption of AI-goods and products. Having this quantitative information can serve as a tool for predicting the diffusion of AI-goods and services in a locality.

**3.3.2.3. Research agenda.** Our ability to take opportunities to learn about the effects of taste, fear and cultural proximity on demand for AI-goods and services depends on our ability to adapt the research toolkit and generate the most enlightening answers. The multidisciplinary Culture-Based Development (CBD) 'toolkit' is a combination of moral philosophy, consumer behaviour, behavioural economics and regional economics and should be used because:

- Moral philosophy is optimal for explicating the mechanisms behind moral sentiments and taste (Smith, 1759).
- Consumer behaviour is the classical field for studying demand using hedonic modelling of preference and choice (Becker, 1996; Scitovsky, 1976).
- Behavioural economics has developed the best knowledge for analyzing fear mechanisms (Kahneman & Tversky, 1979).
- Regional economics is most aware of cultural proximity (Torre, 2008; Tubadji & Nijkamp, 2016) and cultural gravity analysis (see Tubadji & Nijkamp, 2015). Regional economics has also studied the interaction between cultural relativity and fear and has documented its implications on the percolation of ideas and knowledge (Tubadji, Angelis, & Nijkamp, 2016).

Thus, a culture-based analysis is required that combines analyses of consumer behaviour and ethical mechanisms using bounded rationality with regard to the consumption of AI goods and services. It would account for the specificity of cultural tastes and fears in each spatially defined market. Analyses are required that identify tastes for the use of AI goods of services, which explore the role of knowledge about and



level of fear from AI and investigate regional variations in consumers' AI-related preferences and choices.

### 3.4. Science and technology perspective

#### 3.4.1. Perspectives on Artificial Intelligence in the fundamental sciences – Gert Aarts and Biagio Lucini

Research<sup>4</sup> in the fundamental sciences aims to investigate Nature at both the largest and the smallest length scales, at the highest energies, and with complex behaviour emerging from simple underlying laws. In the physical sciences this encapsulates the study of the cosmos, including e.g. dark matter and dark energy, gravitational waves, and black holes, and of elementary particles, including e.g. the Higgs boson, the quark-gluon plasma and new physics beyond the Standard Model. Dynamics at small length scales is determined by the rules of quantum mechanics, rather than classical – Newtonian – mechanics, which introduces an intrinsic indeterminacy in the problem, following the usual probabilistic interpretation. Understanding complex quantum systems, quantum control and quantum information is highly relevant for the paradigm of quantum computing, which surpasses classical computing algorithms in a dramatic fashion and, once available, will make previously incomputable problems solvable. Phase transitions, such as the transition between ice and water, or between magnetic and non-magnetic phases in materials, are manifestations of collective behaviour emerging from simple laws. Order parameters, e.g. the net magnetisation of a material, display the presence or absence of macroscopic order and are connected to the underlying pattern of symmetry breaking, linking phase transitions to the microscopic laws of Nature in a precise way.

The adoption of artificial intelligence (AI) in the fundamental sciences, especially in the form of machine learning (ML), has seen a striking increase in the past 5 years or so (Carleo et al., 2019; Guest, Cranmer, & Whiteson, 2018). While previously a link between ML and the physical sciences existed via statistical mechanics, the methodology developed in physics to analyse large systems with fluctuating degrees of freedom, in recent years the use of ML has exploded and it is now employed in most branches of fundamental science, with increasing success and acceptance.

**3.4.1.1. Challenges.** An overview of ML applications across the fundamental sciences (Carleo et al., 2019) is necessarily incomplete, and most definitely beyond our level of competence. Instead we outline two research areas in which ML and AI are of increasing importance.

The Large Hadron Collider scatters protons and also lead ions at speeds close to the speed of light. Due to both Einstein's theory of relativity and the quantum-mechanical nature of particles at these high energies, many particles are created in these collisions, including rare ones, such as the Higgs boson, which decay almost immediately. Since the number of events at each collision is too large to be stored and investigated afterwards, the selection of "interesting" events, which may contain signals of hitherto undiscovered physics, has to take place in real time, often relying on comparison with simulated data. In the language of ML, this can therefore be seen as a classification problem, with the simulated data providing a labelled training set. The challenges here are manifold (Guest et al., 2018). Searches for new physics may have conflicting demands compared to precision Standard Model measurements, leading to tension in how to handle signal and background. Systematic uncertainties may arise from the use of computer-generated training data and a mismodelling in the simulation. Since the interpretation of the outcome of a neural network analysis is less straightforward than for a more traditional approach, comparisons with theoretical models are more involved. And finally, due to the speed

with which AI evolves, it may be difficult to commit to a certain software framework and embed it in existing analysis packages, before the field has moved on.

As a second example, we consider phases of matter easily accessible in the lab, which are highly relevant for technological applications, such as the storage of digital information and quantum computing. A prime case here is given by materials with permanent magnetisation below a transition temperature; magnetic storage underpins the mechanic hard disc. Superconductors, which provide another example, can conduct electricity without dissipation – and hence energy loss – at very low temperatures, which can be realised in lab conditions. Superconductivity is a broad phenomenon that includes many different realisations. A currently much studied but not yet understood type of superconducting material is the topological superconductor, which could be used to realise quantum gates for quantum computers or, more immediately, provide superconducting cables at room temperature. In this class of materials, superconductivity is encoded robustly in a geometric property of the material itself in an appropriate abstract representation space. The lack of immediate connection between the latter space and the variables measured in experiments makes it very difficult to identify concrete topological superconductors and characterise them. Here unsupervised machine learning can make an impact, by learning material properties and phases of matter from measured data, without providing labelled training sets. Very recently, first steps in this direction have been made by using ML for phase identification in known systems (Carrasquilla & Melko, 2017), which in some cases provides excellent agreement with theoretical expectations (Giannetti, Lucini, & Vadachino, 2018). Extending this to new, not yet completely understood, systems, could provide an improved way to characterise phases in materials such as topological superconductors (Melko, 2017).

**3.4.1.2. Opportunities.** It is easy to identify opportunities where AI and machine learning will benefit the fundamental sciences, analysing the enormous data sets available. Instead, here we will discuss two opportunities where the benefits go both ways, with the knowledge and expertise gathered in the physical sciences yielding a positive impact on AI and ML as well.

In classification problems, ML can be broadly understood as an optimisation problem, in which the parameters of a model function are selected to reproduce as closely as possible the known response on the training set, while avoiding overfitting. In practical applications, this problem requires the use of computational resources. With the growth of the data available and the necessity to obtain a model in the shortest possible timeframe (especially in applications in which inferences are time-critical, e.g. in the financial market or in weather forecasting), the availability of algorithms that are as fast as possible becomes paramount. Computationally power-hungry problems of this type have been well known in Science and Engineering, where they have been addressed with excellent outcomes using a set of techniques collectively known as parallel programming, which allow to distribute the calculation on fast interconnected nodes of a computational ecosystem referred to as a Supercomputer. This by-now mature approach to number crunching, known as high-performance computing (HPC), is offering the possibility to accelerate ML algorithms by orders of magnitude, making tractable problems that previously were not, or shortening the time to solution to a point when a prediction can find timely applicability. Not only is HPC tremendously improving the opportunities provided by ML (Berhofer, 2018), but the synergy goes also in the opposite direction, with disciplines traditionally harnessing HPC benefitting from novel ML approaches. An example is given by calculations in which a set of parameters would need to be optimised. In traditional HPC applications, this problem has been approached with expensive grid searches, sometimes informed by educated guesses. ML offers a radically new approach, which, in addition to the optimisation of existing algorithms (Shanahan, Trewartha, & Detmold, 2018), can

<sup>4</sup> **Acknowledgements** – GA and BL are partly supported by STFC grant ST/P00055X/1. BL is supported by a Royal Society Wolfson Award.

potentially lead us towards the discovery of new and more efficient ones (Liu, Qi, Yang Meng, and Fu, 2017). Indeed, the convergence of ML and HPC, sometimes referred to as high-performance data analytics (HPDA), is one of the most promising and potentially disruptive trends in AI, which both fields can benefit from enormously.

A second opportunity is given by the training and development of the next generation of data scientists. The demand for highly-skilled data scientists has been well publicised and is largely driven by the increasing use of AI and ML across many applications in business, industry and healthcare. Fundamental science plays a dual role in this: while AI and ML are used to interrogate scientific data and enable scientific discoveries, in doing so the involved scientists develop valuable data skills easily transferrable to other disciplines, in-or outside science. This is especially so for the next generation of postgraduate students and early-career researchers, for whom training in AI and ML will be intermixed in the research and skill development from day one. In this context, it is interesting to analyse the outcome of the recent funding call of UKRI – UK Research & Innovation, bringing together the seven UK Research Councils, Innovate UK and Research England – for investment in 10–20 Centres for Doctoral Training (CDTs) focussed on areas relevant to Artificial Intelligence (UKRI, 2018). This call, run across the entire UK research landscape, invited proposals for CDTs to “train the research leaders of the future and equip them with the knowledge, skills and creative approaches the UK requires.” Out of the 84 submitted outline proposals, 37 applicants were invited to develop a full proposal (UKRI, 2018). 16 of these were funded (UKRI, 2019). The main focus areas of these 16 CDTs are listed in the table below. Note that each CDT will train, from October 2019 onwards, at least 50 postgraduate students, in 5 cohorts of 10, via a 4-year cohort-based PhD programme.

Perhaps not unsurprisingly, the majority of the funded CDTs will be active in the areas of healthcare and biomedical research, and of responsible AI, with a focus on accountability, transparency, and the societal and human perspective. Two CDTs are active in the areas of core AI research, language processing, or sustainability and the environment, and one CDT in the development of nano-devices or the creative industries. Only one CDT has a partial focus on fundamental science.<sup>5</sup> It seems therefore that the training opportunities offered by the use of AI and ML in the fundamental sciences are not yet developed enough to convince the main funder in the UK, providing therefore a clear opportunity for the future.

**3.4.1.3. Research agenda.** Fundamental science offers a distinctive but generalisable perspective on future research in AI, as most of the challenges it faces are complex and at the same time deeply rooted in an approach that has evolved from a long tradition in which practical realisations of experiments, theoretical models, and the underlying philosophy are intertwined and entangled in a coherent structure.

For centuries, the route to scientific discovery has followed the scientific method, i.e. observations lead to the formulation of a hypothesis; data is collected to refute or confirm this; subsequent refinements lead to further understanding, culminating in a theory capable of not only explaining all observations so far, but also able to yield new predictions. One may say that artificial intelligence and machine learning represent a new paradigm to do science, inverting the scientific method, by putting data first, especially in the context of unsupervised learning. By inferring patterns from large sets of data in an unbiased manner and building theories to explain these patterns, one skips over the step of testing hypotheses and hence removes bias in the data analysis.

Indeed, most of the current physics experiments are designed with a bias. For instance, the searches for new particles at the Large Hadron

Collider are based on models developed over many years. These models require specific signatures to pick out interesting events. Hence, in a specific search, a trigger will discard all events not containing those signatures. It is natural to ask whether in this way too much information is thrown away, perhaps related to novel interactions and particles that are not part of any currently known model and hence cannot be selected. The question is whether an appropriately setup multi-agent system would be able to select events without a bias, hence leading us towards the discovery of new laws of nature. A similar logic could be applied to most disciplines. Automated science discovery is henceforth a very relevant subject, which should be high in the research agenda.

Automated science discovery cannot be disjoint from interpretability. Outside fundamental science, practitioners of AI and ML are often interested in answers, without associating them necessarily to a complete understanding of how they are obtained, provided that the outcome is reasonable and looks robust. In fundamental science, one would need to go one step further, since from those answers physical laws need to be inferred. This would involve “opening the black-box” (unboxing) that has determined the observed outcome, a long-standing problem in AI, to deduce physical properties from the mathematical expressions the system has worked out for classification or clustering. An example of a successful inference along this line has been provided recently (Wang, 2016). The insights and techniques developed in this area of research would have a wide impact, as unboxing is important for the acceptance of AI approaches in general, with important applications in e.g. the medical and healthcare sector. Finally, a natural follow-up question is whether the inference of the laws governing a phenomenon can be delegated to a second machine, hence going full-circle in the AI-based approach to the fundamental laws of nature.

#### 3.4.2. Science and technology studies – Vassilis Galanos

The present contribution to the joint opinion article comes from my recent investigations and preliminary findings from a Science and Technology Studies (STS) perspective on the social dimensions of AI, robotics, machine learning, and other related meshed and hardly defined concepts. One of my findings is that the clear cut division between challenges, opportunities, and research agenda is quite difficult to achieve. Arguments of challenges reveal opportunities and acknowledgement of opportunities are alternative statements of research agenda points.

The recently appointed House of the Lords’ Select Committee on Artificial Intelligence (HLSCAI, 2018) identifies AI as part of UK’s industrial sector and recommends that: “the Government must understand the need to build public trust and confidence in how to use artificial intelligence, as well as explain the risks” (25). Most challenges identified there are industry-oriented (for example the knowledge transfer from the Academy to industry as well as the data divide between such spinouts, start-ups, small and medium-sized enterprises and large corporations), educational challenges (introducing AI awareness courses at schools), reality distorting challenges due to AI algorithms (e.g. fake news), liability issues between individuals or companies when AI decisions lead to harm. In the same report’s final Appendix, the authors point out that “[c]urrently, EU regulations limited what could be done in this area, but post-Brexit, some attendees felt that there would be opportunities to reassess this” (178).

Previous AI policy documents from the EU, UK, and US (all published in 2016 and of which a good summary and a philosophical review can be found in Cath et al., 2018) have been found to be quite unrealistic in their proposals. The more recent Rathenau Instituut’s report (Van Est, Gerritsen, & Kool, 2017) on Human Rights in the Robot Age: Challenges Arising from the Use of Robotics, Artificial Intelligence, and Virtual and Augmented Reality written for the Parliamentary Assembly of the Council of Europe (PACE) is rather well-informed and examines the relationship between “self-driving cars, care robots, e-coaches, AI that is used for social sorting, judicial applications of AI, and virtual and augmented reality” and “the right to respect for privacy,

<sup>5</sup> Disclaimer: the authors are PI and technical director of CDT #5, see [cdt-aimlac.org](http://cdt-aimlac.org).

human dignity, ownership, safety, responsibility and liability, freedom of expression, prohibition of discrimination, access to justice and the right to a fair trial” while they also recommend the discussion of “two potential novel human rights” in light of AI/robotics-related advancements namely the “right to not be measured, analysed or coached.” These two policy documents are, in my opinion, the ones closest to technical and social reality, although a careful reading pinpoints to the very fact that the great lack of empirical data makes all speculation a challenge in its own right. This is the reason that the most fundamental challenge (and opportunity) to remember is that despite the fact that AI has been very hyped in the last five years, a disciplined turn to specialists and the grounding of research agenda on the basis of technical evidence should be a core priority of any work dealing with the politics and economics of AI.

**3.4.2.1. Challenges.** What are usually presented as AI-related challenges appear to be diversions from rather real-life problems and have little to do with current capabilities of AI. This happens because the very term AI is at the same time ill-defined but also loaded with meaning, expressing hopes and fears ranging from the will to understand intelligence to the consumption by one's progeny (in this case, intelligent robotic overlords) (Szollosy, 2016). AI is specific enough to cause sensationalist alarm and/or excitement, and vague enough for many commentators to interpret it according to their own agenda(s).

This brings us to the very first actual challenge of AI, which has to do with the **problematic aspects of the term**. As argued (Galanos, 2018), both terms “artificial” and “intelligence,” if examined separately have been contested for their rigidity and usefulness. The differentiation between natural and artificial (or nature/nurture, nature/culture, and other similar divides) is impossible to define as either all reality is natural (an outcome of the same nature; naturalist perspective) or all is constructed (interpreted within human brains, a loose constructivist perspective). Such distinctions are posed usually with the conscious or unconscious intention to either “naturalise” (hence justify) certain behaviours (“this is unnatural!”) or to imply human supremacy over nature. Both cases are associated with scenarios of more-than-human or less-than-human, found in horror stories as well as science fiction.

All in all: the “artificial” in AI is in itself a terminological challenge. Similarly, several cognitive scientists, cyberneticians, system theorists, AI/robotics specialists, and sociologists argue that AI is merely indefinable because we do not have any good understanding of the word intelligence (ibid). From empirical data I am gathering at the moment, interviewing specialists with direct (or somewhat direct, given the ambiguity of the term) involvement in AI R&D, I begin to shape the view that specialists are particularly sceptical when it comes to use the term. Most of them tend to “unmask” the term and provide with names of other technologies expressing a syllogism of the type “when people speak about AI, they basically mean x (machine learning, the internet, algorithmic training, deep learning, data science, and so on)”. For the “elderly” AI generation, AI has nothing to do with what is now portrayed as AI, as “true” good ol’ fashioned AI (GOFAI) is based on different techniques and methodologies; also has different purposes (for these and other debates on weak/strong AI see Brooks, 2002; Pickering, 2009; Searle, 1980). This poses very practical challenges in cases where convoluted networks are used to generate deep fake videos; these may be perceived on behalf of the public in tandem with a generalised distrust and less confidence towards (digital) media and the fact that convoluted networks are very easily masked as “AI”. This allows the AI-as-an-enemy narrative to continue in a meshed context of various overlapping technologies. Similarly, robotic hoovers sending data to third parties, generating mirroring effects of targeted advertising, show the data-intense problem which lies behind (and basically allows the existence of) AI fearful hypotheses.

A more tangible challenge is the data basis of what constitutes contemporary commonly perceived AI – and in particular the **data**

**wealth problematic**. In a nutshell: my “AI” will differ to yours if we don't have the same data sources or if you are a company and I am an end-user. (Although this appears as a data ethics/politics argument, given that data is the bread and butter of current AI, I believe it is worth to mention). Sawyer (2008) identifies a gap between data rich and data poor in contemporary cyberinfrastructure. The main reason behind current success in the development of AI is essentially the massive generation of data. Allowing the 1980s pattern recognition-based machine learning algorithms to produce fruitful patterns, a data wealth analysis should become part of top-priority setting when it comes to the economic development of various layers of the market (e.g. data/AI emperors versus start-up AI-based companies) but also with regard to one's personal control of data and the awareness of the various type of uses of their data. A number of recent works have shown (a) the interconnectedness and historical association between automation, data-driven, and AI technologies, and (b) how one of the main problems arising has to do with the generation of new inequalities and the perpetuation of older biases relating to all intersectionality concerns with race, gender, sexual orientation, species, age, and other problematic human categories (e.g. Eubanks, 2018; Buolamwini & Gebru, 2018; Prainsack, 2019). To tie this with above statements, the AI hype and ill-definition diverts focus from problems which should be prioritised instead of policy discussions having to do with robotic liability, and so on which take non-experts' accounts as expert knowledge and science fiction plot devices such as Asimov's laws of robotics at face value (e.g. European Parliament, 2017).

My conviction for this conceptual and terminological burden (which, more than specialists who do not pay attention to definitions as long as things work, does impose problems to other users of the term) is that researchers examining the sensitising concept “AI” should borrow descriptive terminology from similar cases of other studies in complexly defined technologies. One is the understanding of AI as Rorschach, that is, as a psychologist's inkblot shown to the patient, upon which patients project their hopes and fears (borrowed by Turkle's 1981 “Computers as Rorschach” metaphor). The second, more concrete, and with applications in policy, industry, and other relevant social clusters, is the understanding of AI as institutional hybrid, that is, a term quite problematic which needs to be reinterpreted according to the needs of different actors (a lawyer understands AI in a way different to a policymaker, in a way different to a journalist, in way different to a bioinformatics specialist, in a way different to a sci-fi fan, and so on).

The agenda purpose then would be to create typologies of different understandings of AI according to different players/institutions/arenas/other social clusters (this approach borrows largely from Haddow et al.'s 2010 work on xenotransplantation as well as Brown and Michael's 2004 work on biotechnology and how different types of transplants and “risky creatures” were found extremely difficult to fit in various seemingly unconnected areas). A final framework to keep in mind when examining AI terminology (and rhetoric) is Donald MacKenzie's 1999 certainty trough based on an x-y ratio where x represents the proximity to the production of technology and y the degree of certainty. Drawing from missile technologies, MacKenzie proposed that more and the less directly involved one is with the production of technologies. The more uncertain they are, the public knows nothing about it, hence they are uncertain. The developers know way too much, hence they know what might go wrong, so they are uncertain; the in-between intermediaries (promoters, commentators, managers, spokespersons, etc.) with little involvement in the production, wish to appear convincing to the public (buyers), hence their certainty appears to be high (thus, the trough, MacKenzie, 1999). In the case of AI, as I have examined it, this has little to do with promoters as much as it has to do with philosophical, futuristic, and journalistic narratives, perpetuating alarming concerns based on very poor.

This challenge, that is, the intrusion of non-experts to AI debates (or, to be more precise, the development of AI debates by non-experts), I have examined in a recent paper (Galanos, 2018). I believe that the 4-



year period 2014–2018, if examined carefully from a media + policy perspective shows the crystallisation of a third AI hype (after the pre-Lighthill and Alvey Programme ones described in Dwivedi et al., 2015b), established through public commentary by prestigious public personas such as Stephen Hawking, Elon Musk, and Bill Gates who, after using their credentials in domains other than AI, promoted AI doomsday scenarios which found their ways into policy discussions. One of my current hypotheses (to a certain extent verified from various recent initiatives which have not yet taken the form of papers but can be traced in the news or in academic contexts), is that the hype is now entering the trough of Gartner's disillusionment (Linden & Fenn, 2003) and the most important thing to avoid is an upcoming third AI winter. Concerns about a third AI winter have been expressed by AI specialists such as Booch (2015) and now that a general realisation of the fragility of ultraintelligence/singularity/superintelligence types of arguments (for a concise review of the arguments see Eden, Steinhart, Pearce, & Moor, 2012; for their most recent instalment, Bostrom, 2014) has become relatively fashionable (as I often say in my talks “unhype is the new hype”), it is time to strategically intervene and promote justifiable (and yet opportunistic) agendas in AI.

**3.4.2.2. Opportunities.** I find out that clear separation between opportunities and research agenda is relatively difficult to achieve, especially if one follows a relatively sceptical stance towards AI. Hence, I will keep the section on opportunities short and focus more on the agenda below; moreover, opportunities impose expectations and, especially in the case of AI, far-fetched expectations have been harmful. This brings me, however, to the first opportunity (resulting from an earlier challenge).

History of AI shows a repetitive rise and fall pattern of hype and disillusionment; large availability of grants followed by long periods of research support stagnation – this happened because AI specialists, in their attempts at establishing their field have made very brave and overly ambitious (and ambiguous) promises to eventually remain unfulfilled (Crevier, 1993; Fleck, 1982). It seems that the promissory arena has changed and, as it has been shown, in a dangerous fashion for policy (Galanos, 2019), a large amount of negative promissory work has now shaped public opinion through the input of science-related prestigious personas (e.g. Stephen Hawking, Elon Musk) whose undeniable expertise in certain fields allows them to acquire imaginary credentials to become spokespersons of any other “hot” technical/scientific subject. The availability of previous historical sources on the first two rounds of negative effects of promissory work gives the opportunity for a strong basis for (a) investigating in detail the relatively undocumented history of AI past the early/mid-1990s and (b) given the evidence of the current negative effects of non-specialist intrusion, the right to intense boundary work to separate who is entitled to be a spokesperson of AI and who is not.

Research councils, whose active interest in AI is relatively recent (e.g. in the UK, the Eight Great Technologies report was published in 2013) and hence their needs appear modifiable, should be approached by academics investigating AI in an empirical manner and be recommended that certain AI-related challenge funds should be dedicated to (a) explainability/intelligibility of AI/machine learning “black-boxes,” (b) to work evidencing that AI (like every technology) is about augmentation of human skills and not abrupt replacement which is found to be nearly impossible in most cases and (c) to promote in every institution, spinout, and laboratory a framework of strategic foresight (constructing plans and setting goals for a maximum of 3–5 years), instead of circulating abstract expectations looking forward to several decades (cf. the proposals by Van Lente (2012).

### 3.4.2.3. Research agenda for policy and research

- Investigation of the relationship between available data and unrepresented groups – a good case can be made in medical

applications of machine learning when it comes to rare diseases. The development of a strong novel economic framework which will deal with the question of assigning value to data. Whereas traditionally the accumulation of some source tends to decrease its value, with data the opposite seems to appear the more, the better. What are the implications of such a reversal? Is it really a reversal?

- Further, with machine learning as a point of argumentation departure, a consideration of a possible alteration of data protection legislations (e.g. GDPR), will enable people with minimal access to data to be represented.
- Push forward an agenda which will strongly suggest the involvement of social science scholars with expertise in AI-related topics as members of science and technology research committees (e.g. House of Lords, EPSRC).
- As a subsidiary to the above, such social scientists should be responsible for the boundary work between who is entitled to become a witness for such committees and to ensure that the conductors of reports “ask the right questions” instead of biased, (mis)leading, and irrelevant ones.
- Emphatic recommendations for balancing the overwhelming amount of speculative, future-oriented studies on AI (which usually take a technologically deterministic view of the “how AI will change society” type) with empirical-driven research in terms of the history and sociology of AI, separation of actual versus imagined capabilities and challenges.
- Several “elderly” or “traditionalist” AI scholars may agree that the quest of create AI in the early era of the field was mainly the quest to understand intelligence at large (or at least human). Contemporary rhetoric on AI seems to take for granted that humans know what is intelligence and that there should be concerns of the hubris relating to the creation of artificial intelligence. AI (and especially robotics) specialists have to a great extent turned to bio-inspired and non-human-centric approaches to achieve small successes, little by little (e.g. Brooks, 2002) suggesting that we cannot make purist AI since we do not know what is intelligence. A generalised support of the return to the notion that AI should enable the understanding of intelligence should be fruitful for research as it will allow an escape from current misinformed narratives concerned with hubris.
- Based on the recommendations by Winnograd and Flores (1986) and their more recent revisit by Collins (2018), and if the achievement of a more “original” GOFAI (good ol’ fashioned AI) is desired, special grants should be given to natural language processing instead of machine learning. The latter has developed up to a significant degree to assist new findings related to NLP (for example, the intricacies of language as a form of commitment to social obligations and as a social act – or “speech act”) with the assistance of current machine learning algorithms which will enable the understanding of previously unsolvable correlations.
- A constantly reassessed mapping of actors and players shaping the sensitising concept “AI” should be generated in order to gain clearer views of the currently developing system which underpins the AI confusion. How do the not-so-separate-although-sufficiently-distinct clusters of AI developers, AI industrialists, AI commentators and futurists, journalists, end-users of every sort (from Amazon customers to doctors making use of new AI tools), policymakers, science fiction authors and film industries (and so on), relate to each other? How do their relationships change and how does this effect the overall shape of the technology and the public portrayal of AI? An interactive map of such relevant social groups or niches should become available to all such members to raise awareness of the generalised messiness of the meaning of AI.
- The organising of an activist niche which will take mostly into account people with academic expertise in technical/practical AI-related fields, will bring forward and publicise an agenda to discuss the dangers and possible advantages of autonomous weapons. In other words, the opposition between the questions “why make

weapons in the first place?” and “if we are to have weapons, why not make them as accurate as possible?”

### 3.5. Government and public sector perspective

#### 3.5.1. Artificial Intelligence in the public sector – Rony Medaglia

A key area of application of AI technologies is the one of the public sector. The core difference between AI technologies and traditional office automation is that the former do not only support decisions on a pre-programmed if-then logic but, instead, feature learning capabilities (Russell & Norvig, 2016). Given this characteristic, AI presents a special range of opportunities and challenges in a public decision-making context, where environmental variables are constantly changing, and pre-programming cannot account for all possible cases.

AI technologies, such as machine learning, rule-based systems, natural language processing, and speech recognition, when adopted in the public sector, carry potential implications for all aspects of government actions, including the inner workings of government agencies, the relationship between governments and citizens, and the role of governments as regulators (Eggers et al., 2017).

Here we outline the key challenges and opportunities of the specific application of AI in a public sector context, and present a research agenda.

**3.5.1.1. Challenges.** A number of challenges related to AI adoption in the public sector are not unique to AI, but instead overlap with well-documented problems of adoption of any new emergent technology in government. These classic challenges include: the quest for data integration across different organisations, resistance to use by the public sector workforce and citizens alike, and threats of labour substitution (Sun & Medaglia, 2019).

Conversely, we also would like to pinpoint three specific challenges in the adoption in the public sector that are unique, or especially relevant, to AI: algorithmic bias, algorithm opacity, and filter bubbles.

The first challenge concerns algorithmic bias. AI-based algorithms are increasingly being experimented by governments to introduce efficiencies in the large scale customisation of public services, a type of task that draws on citizen profiling (Janssen & Kuk, 2016b). Examples of such applications include public hospitals using machine learning algorithms to predict virus outbreaks (Mitchell, 2019); analytics tool used to predict hotspots of crime (Goldsmith & Crawford, 2014) and high risk youth (Chandler, Levitt, & List, 2011); and AI systems used to target health inspections in restaurant businesses (Kang, Kuznetsova, Luca, & Choi, 2013).

While the ability of AI applications to recognise patterns can be beneficial to segment populations for e.g., welfare service provision or addressing anti-social behaviour, it can also amplify discriminatory biases that are already present in human-led assessments: predictive algorithms, in fact, can favour groups that are better represented in the training data (Barocas & Selbst, 2016). Algorithms can thus lead to systematic and unfair treatment of citizens based on social biases of gender, race, sexuality, and ethnicity – an outcome which is in direct conflict with the mission of governments of unbiased treatment of citizens under the rule of law.

The second challenge concerns algorithm opacity. The increasing complexity of AI systems, such as machine learning and neural networks, reduces the capability of human operators to trace outputs back to specific inputs, making it potentially impossible to clearly account for specific AI-driven outcomes. The wider consequences of this phenomenon have been referred to as creating a “black box society” (Pasquale, 2015), and have profound implications for governments which, by definition, are bound to citizen expectations of transparency and accountability. A clear example can be found by looking at how digital systems impact the work of street level bureaucracies. While, on the one hand, automated decision-making has the potential to improve fairness by reducing the discretion of public service operators (Busch &

Henriksen, 2018), on the other hand AI systems can remove public servants from the duty of accountability, exacerbating the phenomenon where citizens are faced with impotence in front of “the computer says no” responses (Wihlborg, Larsson, & Hedström, 2016). The opacity of mechanisms in AI-supported decisions poses challenges not only in the ethical responsibility and legal liability dimensions – who is responsible for a damage to a citizen stakeholder, if the decision has been outsourced to an AI application? – but also to the wider fundamental issue of political accountability of public governance.

The third challenge is associated with the creation AI-enabled filter bubbles in the public sphere. AI-enabled algorithms have proven tremendously effective at micro-targeting content and at fostering a booming constellation of groups of like-minded actors in the public space, such as social media platforms (Sunstein, 2017). This challenge affects the context in which public governance is exercised – that is the sphere of public opinion formation at large and thus, indirectly, the ability of government to both be seen as legitimate by citizens, and to formulate policy actions that draw on a perceived common good. The ability of algorithms to provide personalised content by filtering out inputs that do not match pre-existing user preferences (in e.g., news, entertainment, political discourse) is potentially bringing about societal fragmentation, polarisation, and radicalisation, with the creation of digital echo chambers (Medaglia & Zhu, 2017). Governments that fail to mitigate such disaggregating forces, enabled by AI systems, will potentially lose the capability to be perceived as legitimate and to formulate policy actions that can be met by sufficient public opinion support.

**3.5.1.2. Opportunities.** The introduction of AI in the action of government comes with a wide range of unique opportunities. While many of them start to be highlighted in a booming number of viewpoints on AI in the public sector (Desouza, 2018; Duan et al., 2019), we focus here on two specific ones that we consider outstanding: relieving cognitive resources of public workers, and fostering citizen trust.

First, AI applications, such as rules-based systems, speech recognition, machine translation, computer vision, machine learning, robotics, and natural language processing, have the potential to free up precious cognitive resources of public workers, which can then be allocated to tasks of higher added value (Eggers et al., 2017). This reallocation allows government to focus scarce resources on tasks at which human workers perform better than machines, such as problems solving activities that require empathy, creativity, and innovation.

Second, AI applications have the potential of fostering citizen trust. The other side of introducing “digital discretion” in the work of street level bureaucrats by AI systems is that unfair, inefficient, or distorted provision of government services can be potentially reduced, thus increasing citizen trust towards government. The introduction of traditional digital government initiatives has already been documented as helping reduce public servant corruption (Bertot, Jaeger, & Grimes, 2010): AI systems can bring this trend further, provided that governments put great care in ensuring that the adoption of AI is included in a context of dialogue with citizens, and towards counteracting the increasing distrust towards governments. Both the ability of AI systems to micro-target policy recipients (which allows governments to implement much more fine-grained policies), and to ensure real-time, flexible rule-based action by street level bureaucrats (which reduces arbitrariness and citizen divides in service provision), can potentially enhance citizen trust towards governments.

**3.5.1.3. Research agenda.** The layered nature of the potential disruptions in the introduction of AI in the public sector calls for a renewed research agenda, and new theorisation efforts (von Krogh, 2018). Here we identify three key areas of research, to be prioritised in the near future.

First, there is a need to unpack the impacts on the public sector



workforce of delegating decision-making to AI. Besides highlighting the classic threat of labour substitution, what the introduction of AI systems calls for is research on the nature and the mechanisms of transformation in the public workforce. Automated systems can undermine worker motivation, cause alienation, and reduce satisfaction, productivity, and innovation (Moniz, 2013). Research questions worth investigating will thus be: what are the motivational and psychological impacts of introducing AI as a “digital colleague” on the public workforce? How are inter- and intra-organisational dynamics of power in public agencies shaped by the introduction of AI?

Second, there is a need to better understand the dynamics in the attribution of meaning to AI-supported public decision-making. As AI applications tend to introduce opacity, and reduce the ability of non-experts to audit the mechanisms that lead to decision outputs, we need to unpack the novel sensemaking processes enacted by government workers and citizens alike, when facing AI applications. Examples of research questions include: how do policy makers frame and legitimise AI-supported solutions? How do citizens perceive the role of AI in policy making? How is agency attribution formulated and negotiated between different stakeholders in the government sphere?

Third, there is a design research challenge to tackle the issue of the potential opacity of AI applications. The nature of AI algorithms seems to suggest that their transparency, traceability, and explainability are inversely proportional to their complexity. While this might be the case, there is no reason to consider such characteristics as immune from mitigation strategies in the design and management of AI applications. Research question related to this challenge include: how to design algorithms that enable explanation? How to design evaluation frameworks that avoid discrimination?

The introduction of AI in the public sector opens up new scenarios for practitioners and researchers alike. Being able to understand and act on these scenarios becomes now of utmost importance.

### 3.5.2. AI for SMEs and public sector organisations – Sujeet Sharma and JB Singh

AI technology although gradually developed in the past several decades, has accelerated shown more in the past number of years due to promising developments in machine learning algorithms, rise of big data and low cost processing power due to the advent of cloud computing. Although AI comprises of a set of technologies such as machine learning, deep neural networks, natural language processing, robotics etc., in simpler terms it can be defined as an advanced prediction technology (Agrawal, Gans, & Goldfarb, 2017). In this sense, AI technologies can find patterns in large amount of data and provide predictive outcome for the new similar instances. The well documented AI applications such as cancer detection in health care and fraud detection in financial industries are promising. AI applications in various industries and activities such as manufacturing, human resources, and sales and promotion are growing, however such applications are currently limited to mainly larger business enterprises.

**3.5.2.1. Opportunities.** Artificial Intelligence applications have potential for SMEs, public/government organisations and also not-for-profit enterprises where the potential has not yet significantly explored. The potential of AI in SMEs are particularly in automation of various tasks with decision making components such as in the functions of finance and customer services. For example, AI applications could help SMEs in matching customer invoices with received payments, AI chatbots could help these enterprises answering customer's simple requests. These are some easy to implement applications where AI could improve the efficiency of SMEs. Unlike SMEs, public sector and government organisations generate lot of data through their processes and hence more potential for application of AI technologies. In developing countries particularly, payments of the welfare schemes to the eligible citizens is paramount in functioning of the state. One of the problem in developing countries is the leakage due to corruption and

appropriation of such schemes by the elites or the well-connected set of people. AI could help in identifying the target citizens for such welfare schemes and payments. Another promising area of AI applications could be in judiciary in developing countries where cases are pending from decades due to availability of limited resources. AI could help in deciding on the bail hearings in courts as machine learning technologies are now robust for such applications (Mullainathan & Spiess, 2017) and may deliver decision which might not only be quick but also more accurate.

Similarly, AI opportunities also exist in the not-for-profit enterprises, one of the fine examples is Akshaya Patra in India,<sup>6</sup> which runs world's largest mid-day meal programme serving wholesome food to the children of government and government aided schools which has the aim of reducing malnutrition and facilitating the right to education of socio-economically disadvantaged children. Here, AI could be used to accurately forecast the demand of the meals for schools based on the data of student's attendance records and hence greatly minimising the waste of food (Raval, 2018).

New technology adoption in SMEs and public sector enterprises generally followed by the adoption in large enterprises when such technology becomes stable and affordable. However, in case of AI applications cost may not be the bigger issue as the supporting software programmes are increasingly also available as open source.<sup>7</sup> SMEs and public sector organisations need to understand the capabilities of AI technologies and should work towards appropriating these capabilities for solving existing business concerns. In the current scenario, most of the popular AI applications in media reporting are of large and information technology intensive organisations.

**3.5.2.2. Challenges.** Large private sector companies such as Google, Facebook etc. are adopting artificial intelligence enabled tools to obtain competitive and strategic advantage in the digital marketplace. These companies maintain enough information resources in terms of information technology assets and capabilities to exploit data for better decision-making. However, SMEs and public sector companies may face some challenges in leveraging artificial intelligence enabled tools. There are many challenges in adopting AI based tools ranging from data quality, privacy/security risks and to the shortage of qualified AI experts. In this section, we attempt to describe some of the major challenges faced by SMEs and public sector companies.

**Data quality:** Data quality can be thought of as the fitness of data to obtain actionable insights using appropriate analytical tools (Lee, 2017). Data noise, data heterogeneity, imbalanced data, data discretisation are some of the prominent reasons of low quality data. In the popular epic poem Rime of the Ancient Mariner, author stated that, “Water, water, everywhere, nor any a drop to drink.” The usefulness of the data depends on the quality of the available data in the companies' warehouses (Hazen, Boone, Ezell, & Jones-Farmer, 2014). Low quality data lead to the poor decision-making and as a result loss in businesses. The cost of the low quality data may lead to loss of 8% to 12% of the revenue in an organisation and may translate in the loss of billions of dollars in a year (Dey & Kumar, 2010). As most of the data collected over a couple of years is unstructured and amassed from multiple sources, the overall quality of collected such data is assumed to be low in the SMEs and public sector companies. If low quality data is used to train AI enabled tools, it will lead to disaster. In public sector organisations, there are rarely available data standards to collect and store data which results in low quality. Therefore, data quality is one of the key challenges in the adoption of AI enabled machines and becomes severe in SMEs and public sector enterprises.

**Privacy/security risk:** Privacy and security are key challenges in adopting AI enabled tools in any organisational settings. These

<sup>6</sup> <https://www.akshayapatra.org/about-us>.

<sup>7</sup> <https://dzone.com/articles/how-open-source-software-drives-iot-and-ai>.

challenges become severe in case of SMEs and public sector organisations due to availability of limited resources. In general, SMEs work under severe constraints of ICT resources and their primary objective is to buy and sell products. In this process, SMEs generate reasonably good amount of data related to product and users. There is a challenge to maintain privacy and security of such useful data. In public sector enterprises, there is a huge amount of personal data being generated during citizen centric services rendered by government agencies. This huge amount of data is vulnerable to data theft or data manipulation as ICT regulations are quite weak in many developing countries. In addition, privacy is a major contributor of legal and ethical concerns raised by the rapid growth of AI enabled products in past couple of years. Duan et al. (2019) discussed that ethical and legal issues are major challenges of AI enabled services.

**Shortage of qualified AI experts:** In SMEs and public sector companies, leaders are attempting to employ AI to see positive impacts on the business outcomes and hunting for AI experts to transform their vision into reality. Bernard Marr (2018) reported that there is a requirement of one million AI experts worldwide but available AI experts are about 300,000. Bernard Marr (2018) further argued that the shortage of AI experts is due to the mismatch between skills taught in an academic environment and skills expected to keep pace with new AI technologies. In addition, there is a well-established relationship between artificial intelligence and data science. In fact, AI is considered as a tool to data science that provides actionable insights to a particular problem. In a recent study, The Economist Intelligence unit (2018) conducted a survey of 400 senior executive working on the transformative potential of artificial intelligence in private and public sectors across eight prominent markets including USA, UK, France among others. This report reveals that “talent and skills” is one of the business’ top strategic challenges in the current scenario. Davenport and Patil (2012) claimed that data scientist is going to be the sexiest job in the 21st century. Vesset et al. (2015) reported that there will be shortage of data science experts and will grow at a compound annual rate of 23% by next couple of years. In general, shortage of qualified AI experts is another important challenge but it becomes critical in case of SMEs and public sector enterprises.

**3.5.2.3. Research agenda.** The above opportunities and challenges discussion provides further avenues for research directions. Each of the points discussed above need greater attention from scholar to conduct in-depth research studies. In this section, we propose the following research directions

- Given the constraints in terms of skilled talent, data quality, and privacy and security, there is a need to understand whether SMEs and public sector organisation should adopt the strategies adopted by large organisation or do they need to formulate new AI strategies?
- Another possible research direction is to assess readiness of SMEs and public sector organisations for AI applications.
- Finally, there is also need to understand and explore the impact of security and privacy risk in adopting AI applications in SMEs and public sector organisations.

**3.5.3. Public policy challenges of Artificial Intelligence (AI): a new framework and scorecard for policy makers and governments – Santosh K Misra<sup>8</sup>**

The emergence of AI as a potentially disruptive technology has posed new challenges for policy formulation in the 21st century. AI can be thought of as a computational algorithm that is capable of learning and identifying patterns in a given voluminous data set and then able to

apply this learning to new unseen set of data in order to make an autonomous decision without any human supervision (Negnevitsky, 2011; Stone, 2016). Today, AI applications are touching human lives in every sphere – self driving cars, medical diagnostics, drug discoveries, law enforcement, military, space, education, governance and elderly care are just a few example. A report of McKinsey Global Institute estimates AI contribution to global economy at US \$13 trillion by 2030 (Bughin et al., 2018). The same report estimates that about 70% companies would be using AI by 2030.

Massive portable computational power available ubiquitously around the globe is the new reality. This compute power coupled with thousands of open source AI modules available on platforms such as github, has transformed AI systems into a commodity which can be bought and sold ‘off-the-shelf’ across the globe. This has made the risk management of an AI system very complex (Scherer, 2016). This has also raised some very challenging issues for the Governments and needs a well thought out systematic Public Policy response. For the policy makers one of the key challenges lies in staying ahead of the technology curve and in being able to identify new technological disruptions taking shape. The goal of new Public Policy should be to allow harnessing the power of AI for public good while keeping it safe and ethically compatible with human values. The AI systems of future, being capable of autonomous decision making – which in areas like law enforcement or healthcare may interfere with right to life or right to freedom of a human being – must be designed to be compatible with our social values, ethics, fairness, equity and our idea of accountability. This is critical for survival of a free human society and it cannot be left to the wisdom of private capital, which however well meaning, is likely to put a premium on the bottom line numbers over everything else. Interestingly, the loudest call for regulation by Government is emanating from the tech leadership itself. Researchers have called for creating National bodies for oversight on AI and Algorithms (Gaon & Stedman, 2019; Shneiderman, 2016). Traditional public policy and regulatory responses such as licensing, R&D oversight and tort are not suitable for AI, because of the discreet, diffused and opaque nature of AI (Bleicher, 2017; Scherer, 2016).

Governments, unlike private sector, have twin roles in the adoption of AI –

- **As a user of AI** – to better deliver the services to citizens, to improve efficiency, to cut down waste and to optimally allocate resources
- **As a regulator of AI** – keeping the technology benign and oriented towards improving the lives of its citizens. It must lay down policies and framework to ensure all usage of AI is fully compatible with human values, and must ensure that the use of this technology is inclusive and it does not leave anyone behind.

**3.5.3.1. Opportunities for AI in governance.** Governments are responsible for delivering a large number of transaction, licensing and regulatory services to citizens and companies, and are going to be amongst the biggest adopters of AI. The reason for this is not very hard to fathom. Governments world over, invariably grapple with following common problems:

- Perpetually short on resources – Governments everywhere need more resources than they have at their disposal. This requires an optimal allocation of resources which is a highly complex task even for the smallest of the Governments.
- Scale of operation – Scale of operation of Governments is huge and with mandatory burden of centralised recordkeeping of individual transactions for audits, courts, Right to Information etc., it quickly becomes a gigantic and complex task. The extra recordkeeping need makes the processes cumbersome and results in unavoidably complex systems which lead to delays, adversely affecting the quality of service delivery.
- Standardisation – one size fits all approach – To address the first two

<sup>8</sup> The views expressed by the author are his own and do not reflect the views of Government of Tamil Nadu, India.

problems most Governments have resorted to standardisation of systems and processes. While this is great for record keeping and audit trails, it invariably makes the processes and forms significantly more cumbersome for the citizens, and adds more load to the already overloaded Government delivery systems.

Governments could really transform their service offerings by using AI to address all the three issues. Services can be “tailor made to Individual needs” – a 100% customised services to every citizens. The AI can deliver the following benefits –

- **“Smart service”** – Efficient delivery can cut down on time and cost of service delivery and can improve the processes. Autonomous systems and Intelligent Chatbots can continue delivering services 24 × 7, without any ‘off’ days. It would reduce cost and time both for the Government and the citizens. Governments can free up precious human resource from repetitive work and can re-deploy them more meaningfully.
- **“Intelligent Adaptive Forms”** – 100% Customised services to every individual – tailor made forms can be generated with help of AI for every individual based on her/his age, gender, literacy level, special needs, and eligibility. An example is income tax return forms – can it be customised for every individual, instead of citizens expected to write ‘not applicable’ at scores of places. It can easily pull out relevant details from the existing databases (previous year’s returns for example) and autofill most of the fields for the citizen, making it easy and less time consuming. Ideally, any form filling for Government should be just a question answer based mechanism, where just by answering a few questions the required form can be auto generated for the citizen.
- **“Predictive service delivery”** – Using AI and data analytics, Governments can take a big leap forward in service delivery and can start ‘predictive service’ delivery – where the citizens do not need to explicitly apply for every service or benefit needed by them, instead the services or the benefits get delivered to them automatically when they need them. For example, Government can sanction scholarship for a student based on the data it already has (her education performance, socially disadvantaged status, stream of education, parental income etc.) and the student just needs to give a consent for accepting the scholarship over the short messaging service or an automated voice platform. Similarly, a senior citizen just needs to give his/her consent over mobile phone to start getting the social security pension she/he is eligible for. A farmer, a small trader, a micro enterprise or a skilled service provider can get his/her incentives or benefits due from the Government without having to fill a form or going to any Government office.

**3.5.3.2. Public policy challenges of AI.** The Public Policy is facing unprecedented uncertainty and challenges in this dynamic world of AI. Everyday a new application based on AI is invented and unleashed onto the human society. The velocity, and scale of impact of AI is so high that it rarely gives the public policy practitioners sufficient time to respond. Public Policy, by definition, needs to put in place regulations against a possible future development which could be detrimental to human values. This creates an interesting tension between the need to predict AI impact and inability to draw boundaries around this highly dynamic technology. Klaus Schwab, while highlighting the Governance challenges due to AI or what he calls the ‘Fourth Industrial Revolution’, writes:

“Agile governance is the response. Many of the technological advances we currently see are not properly accounted for in the current regulatory framework and might even disrupt the social contract that governments have established with their citizens. Agile governance means that regulators must find ways to adapt continuously to a new, fast-changing environment by reinventing themselves to understand better what they are regulating.” (Klaus Schwab, 2016).

There have been warnings against over-regulating AI, lest it should strangle its development and make future advances either impossible or too expensive (Adam, O’Sullivan, & Russell, 2017).

The Stanford “Report on Life in 2030” has the following three recommendations for the Governments (Stone, 2016):

- Define a path towards accruing technical expertise in AI at all levels of government. Effective governance requires more experts who understand and can analyse the interactions between AI technologies, programmatic objectives, and overall societal values.
- Remove the perceived and actual impediments to research on the fairness, security, privacy, and social impacts of AI systems.
- Increase public and private funding for interdisciplinary studies of the societal impacts of AI.

Intel corporation in its white paper on Public Policy opportunities in AI flags privacy (two components – Fair Information Practice Principle and Privacy by design), accountability, fairness and human employment as key areas of concern for AI policy (Intel, 2017). Germany has identified transparency, privacy and ethics as three critical challenges for AI development and adoption (Harhoff et al., 2018). In UK, the British Standards Institute in its draft proposal on “Information Technology – Artificial Intelligence – Risk Management Standards” has identified transparency, verifiability, controllability, explainability, robustness, resiliency, reliability, accuracy, safety, security and privacy as important parameters for certification (BSI, 2019). The famous Asilomar AI principles list out – safety, failure transparency, judicial transparency, responsibility, value alignment, human values, personal privacy, liberty & privacy, shared benefit, shared prosperity, human control, non-subversion, arms race and strict control of recursive self improvement AI – as key challenges of AI implementation (Future of Life Institute, 2017). Calo in his paper has identified – justice and equity, use of force, safety and certification, privacy, power, taxation and displacement of labour as the key challenges of AI (Calo, 2017). Japan’s METI (Ministry of Economy, Trade and Industry) has listed employment, skilling, database protection, changes needed in laws and global collaboration as main policy challenges for AI (METI - Ministry of Economy, 2016). Canada is being advised by its researchers to focus on trust, transparency and accountability as prime AI policy challenges (Gao & Stedman, 2019). India’s National Strategy for AI prioritises – fairness, transparency, privacy and security as the key challenges of AI over the rest (Niti Aayog, 2018).

It is evident that there is a wide variation in identifying the key public policy challenges of AI. There is a need to unify these approaches and create a unified practical framework for ‘Public Policy Challenges of AI’. This framework must cover all the critical challenges of AI and yet keep the set relatively small to make it tractable and implementable.

Attempts to understand the AI impact on society through literature review (Wirtz et al., 2018) have resulted in classifying the AI impact in public sector in 4 broad areas of – AI & Technology, AI & Society, AI & Law and AI & Ethics. While this approach is a good beginning point for trying to understand the impact of AI on society, what is needed by the public policy practitioners is a toolkit for objectively analyzing an AI for public use proposal.

**3.5.3.3. A new framework for public policy for AI: TAM-DEF.** So far, there has not been any attempt in devising a comprehensive Public Policy framework for AI which would guide and enable the public policy practitioners and Governments in making a decision on using a particular AI system. The TAM-DEF framework proposed here does precisely that. It creates a framework on which Governments can objectively test any AI system before launching it for public use. It provides a systematic framework for the questions Governments must ask before using any AI system. It also provides a DEEP-MAX scorecard mechanism for making an objective decision about intended AI use. In short, it provides a toolkit which can help public policy practitioners in

assessing the safety and social desirability of any AI system. The TAM-DEF framework (Transparency & audit, Accountability & Legal issues, built in Misuse protection, Ethics, Fairness & Equity, Digital Divide & Data Deficit) identifies six key AI Public Policy challenges as following –

- Ethics
- Transparency & Audit
- Accountability & Legal issues
- Fairness & Equity
- Misuse protection
- Digital divide & Data deficit

Governments and regulators would need to address each of these six challenges before rolling out any AI solution for public use. They would need to ensure that each public AI system is minutely examined under the set of six challenges provided above. Only when an AI system is clearly understood on all the above six parameters, it should be cleared for public use.

It is important to highlight that the six dimensions of Public Policy challenges mentioned above, are not watertight compartments (Fig. 2), instead they tend to be a diffused continuum, which Governments must tackle for making AI safe and useful for its people. Now let us examine each of the six challenges in detail:

- **Ethics:** Ethics for machines has been an area of immense interest for the researchers. However, defining ethics for machines has proven to be difficult, and to make it computable has been even more difficult (Anderson & Leigh 2007). To tackle this, TAM-DEF framework treats Ethics purely from AI perspective and divides it in two sub-components – (i) Privacy and Data protection and (ii) Human and Environmental values. Both these dimensions of ethics are critical for keeping AI systems safe for the human society.

○ **Privacy–Data Protection:** Privacy is possibly the top most concern while using AI systems. User's intimate and highly granular data is likely to be stored and shared across the AI network (for example a person's location for the day based on face recognition and CCTV feeds, food habits, shopping preferences, movies, music etc.). The AI systems must ensure that this data remains protected and Governments need to make strong data protection laws to enforce it.

○ **Human and Environmental Values:** Any AI system has to conform to human value system and the policy makers need to ask – Has the AI system been sensitised to human values like respect, dignity, fairness, kindness, compassion, equity or not? Does the system know that it has a preferential duty towards children, elderly, pregnant women, sick and the vulnerable?

An important aspect which needs to be built into AI systems is the overall cost of their decisions on the society. An AI system for example designed to find a particular mineral let us say, would be highly optimised to obtain it. It would try to maximise its output of that mineral but would it be capable of assessing the collateral damage to the environment its strategy is causing, or would it be able to account for pollution externalities created by it. To be able to do this, AI systems should not be optimised uni-dimensionally but need to be trained to factor in their 'world' or environment within which they operate.

- **Transparency and audit:** In the visible future many of the AI based autonomous systems (robots) would be regularly interacting with humans in fields like finance, education, healthcare, transportation, elderly care etc. The technology providers must explain the decision making process to the user so that the AI system doesn't remain a black box to them (The Economist Intelligence Unit, 2016). Moreover, there is a legal need to explain the decision taken by such systems in case of litigation. These AI systems must provide an audit trail of decisions made not only to meet the legal needs but also for us to learn and make improvements over past decisions.

- **Digital divide and 'data-deficit':** Since the entire AI revolution has data at its foundation, there is a real danger of societies with poorer access to information technology, internet and digitisation being left behind. Informed citizens would tend to gain disproportionately in this data driven revolution. Countries and Governments having good quality granular data are going to derive the maximum benefit out of this disruption. Countries where the data is of poor quality or of poor granularity would be left behind in harnessing the power of AI to improve lives of its citizens. There is threat that this technology would adversely affect communities which are poorer in data. Unfortunately it is the low-resource communities in developing countries which would be hit by this data-deficit because they are the ones who never had the resources to invest in data collection and collation.

Another challenge that emerges from this technology is the skewed power distribution between digital haves and have-nots. Only those who have the ability, knowledge and resources needed to connect to online data driven systems would be heard. The voices of others may not get registered in the system.

- **Fairness & equity:** As discussed earlier, AI can, and AI would disrupt social order and hierarchy as we know them today. It can create new social paradigms, which if left uncared for, can severely damage the social fabric and expose people lower in the bargaining hierarchy with a real threat of exploitation and unfair treatment. It would lead to commoditisation of human labour and could chip away at the human dignity.

An AI system designed with equity as a priority would ensure that no one gets left behind in this world. While 'equity' may have some overlap with 'digital divide' (digital inequity) listed above, the concept of equity here covers a much wider range, of which 'digital equity' is just one part.

Another key need for autonomous systems is fairness. They must be 'trained' in human values and they must not exhibit any gender or racial bias and they must be designed to stay away from 'social profiling' (especially in law enforcement, fraud detection, crime prevention areas). The recent reports questioning the neutrality of AI systems used by Police to identify crime prone individuals has brought this issue out in sharp focus (Dan Robitzski, 2018).

AI systems designed must comply with 'free of bias' norm to prevent stereotyping. In MIT Technology Review of Feb 2018, Timnit Gebru highlights the pitfalls of AI designed without diversity incorporated in its base, "If we don't have diversity in our set of researchers, we are not going to address problems that are faced by the majority of people in the world. When problems don't affect us, we don't think they're that important, and we might not even know what these

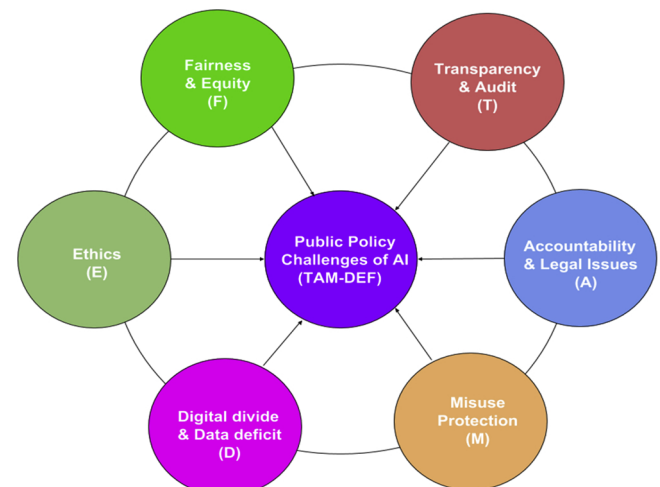


Fig. 2. TAM-DEF framework for public policy challenges of AI.



problems are, because we're not interacting with the people who are experiencing them" (Snow, 2018).

- **Accountability & legal issues:** Without artificial intelligence any system designed by human is only a machine under the control of the operator. Therefore there never is an issue of who is accountable. Almost all civil and criminal liabilities laws of the world, fairly unanimously attribute accountability to the operator, owner and manufacturer of the machine in varying degrees depending upon the facts of the case (Nambu, 2016). However, once machines get equipped with AI and take autonomous decisions, the accountability question becomes very hard to answer. More so when the algorithm used for decision making is sometimes even unknown to the designer himself. AI machines are capable of inventing superior ways of accomplishing the task given, using a purely unintended route. This can have serious implications for the society. The famous case of Facebook AI project where two robots started talking to each other, in an invented language to accomplish a negotiation task they were given, is a sharp reminder of unintended consequences which can emerge. The robots were taught to converse using Natural Language Processing (NLP) but they invented a more efficient communication strategy which looked like gibberish to humans. A snapshot of their conversation as reported in 'The Independent' (Griffin, 2017):

"Bob: i can i i everything else....."

Alice: balls have zero to me to me to me to me to me to me to me to me to me to

Bob: you i everything else....."

Alice: balls have a ball to me to me to me to me to me to me to me to me to me to

Bob: i i can i i i everything else....."

This ability to learn on their own using what is known as reinforcement learning, can have highly unpredictable consequences. One of the leading coalitions for AI, OpenAI has recently created an AI driven text generator called GPT2. The accuracy and creativity of the GPT2 is so high that OpenAI has this to say about its code release –

"Due to concerns about large language models being used to generate deceptive, biased, or abusive language at scale, we are only releasing a much smaller version of GPT-2 along with sampling code." (<https://openai.com/blog/better-language-models/> accessed 23Mar2019)

- **Misuse protection:** This possibility is the toughest of all six questions. How do we fool-proof every new technology so as to prevent it from being twisted for achieving destructive goals. A case in point – the Internet. How internet proliferated across the globe benefitting billions but also carried along with it a wave of cybercrime, malware, viruses and games like 'blue-whale' which resulted in loss of innocent lives of teens around the world. A stark reminder of how destructive the potential misuse of AI technology can be, is the case of FBI agents monitoring a hostage situation related to organised crime in the winter of 2017 in US. The criminals using a swarm of drones managed to force the FBI agents out from their location and they live streamed the video to their gang leader on youtube (Tangemann, 2018). AI systems can also be used by dictatorial Governments for extending their unlawful regimes and suppressing freedom.

**3.5.3.4. Setting safety standards under TAM-DEF.** A few scholars have argued for keeping a tight control over every new technology and not releasing it to the public till its potential misuse is identified and substantially mitigated (Narayanan, 2013). This is likely to remain just a wishful thinking, the pace of new technology development is too rapid to even try and leash them. However, building safeguards by appropriate regulation is certainly what the Governments of the world need to be doing, and preferably doing it collectively.

This is where the role of public policy becomes central.

Governments world over need to agree on a set of standards which every AI rollout must be rated against. An AI system with ratings, would make the user aware of the possible handicaps of the system s/he is using. While the global agreement on the standards may be difficult to negotiate, I believe our purpose would be greatly served even if the national Governments create their own standards under the TAM-DEF framework.

**3.5.3.5. Overlaps in TAM-DEF framework.** Given the complexity of the problem of drawing an outer contour for all possible AI challenges from a public policy perspective, the proposed TAM-DEF framework is an attempt to find a reasonably comprehensive, practical and tractable framework on which any AI systems can be examined for public safety and social desirability before roll out.

The six challenges of the TAM-DEF framework, even though largely independent, are not mutually exclusive. For example, Digital divide in some sense can be linked to the Equity and fairness, but it is important to understand why they are treated separately. Digital Divide is treated separate from Equity, to accentuate the fact that over half the world population has no access to the internet (Source: World Bank, <https://data.worldbank.org/indicator/it.net.user.zs> accessed 14Apr2019) and there are large communities which do not have any data to train any AI system. While the Equity accounts for the traditional meaning of the word, the prevalent Digital divide in the world is too huge to club it under the equity frame. It needs to be treated separately especially when we are talking about a purely digital and data driven technology like AI.

**3.5.3.6. Implementing the TAM-DEF framework – AI Standardisation, DEEP-MAX Scorecard and use of Blockchain for Transparency and Trust.** To handle the six AI challenges mentioned above, a four pronged strategy is proposed for the public policy practitioners and Governments.

First, since the AI systems have a global reach – they are developed in one part of the world and deployed in another – there is a need for a global alliance for AI standardisation and rating.

Second, an objective scorecard (called DEEP-MAX, described below) based on the TAM-DEF framework is proposed, which, with suitably designed test data sets can reliably produce a safety and social desirability score for a given AI system by testing it against each of the seven DEEP-MAX parameters.

Third, use of Blockchains in training, testing and misuse protection of AI Systems could be a reliable mechanism for verifying a safe and socially desirable AI solution. An AI Certification Transparency & Scorecard Blockchain (ACTS-B) can integrate the information about the dataset which was used for training an AI system and it can track whether the training dataset met important criteria like diversity, equity etc. Similarly, the ACTSB would also carry the 7 scores from the DEEP-MAX Scorecard for a given AI system. ACTSB should be a universal publically viewable Blockchain. This would create a transparent mechanism for rating and understanding of AI solutions before putting into use.

Fourth, since many of the AI systems are self learning, the DEEP-MAX scores which would ship-out with each AI module, may no longer be valid after sometime and they would need to be updated. A periodic update of the DEEP-MAX scorecard would need to be mandated for all AI systems deployed for public use. The needed periodicity of update would have to be established based upon the nature of AI use case class.

**AI standardisation and rating:** Like the ICANN for the internet, there is an urgent need for setting up an independent and transparent Global Alliance for AI Standardisation and Rating which should regulate the AI development, testing and rating system for every AI module or system being created. However unlike ICANN, this global alliance must be made a truly democratic international alliance of Governments. Since AI systems developed in one country are likely to be deployed across the world, transparent and uniform standards would provide the

users or developers adequate clarity and confidence in rolling out AI systems. It would also remove civil and criminal liability uncertainty which a company would otherwise face while rolling out its AI products across different legal systems of the world.

The tasks before such a Global Alliance for AI Standardisation and Rating would be:

- Defining privacy standards to be met by all AI systems
- Defining Ethical boundaries for all AI development
- Defining civil and criminal liability of AI systems and a mechanism to deal with them
- Define audit standards to help explain the decision taken by autonomous AI systems

**DEEP-MAX scorecard:** The DEEP-MAX Scorecard proposed here, is a transparent point based rating system for an AI systems on 7 key parameters of Diversity, Equity, Ethics, Privacy and Data protection, Misuse protection, Audit and Transparency, Digital divide and Data deficit (Cross geography and cross society applicability and performance of AI system). Users, System Integrators, or Government Departments designing, developing, or using any AI system can just look at the DEEP-MAX scores of all the individual AI components (which are likely to have been picked off-the-shelf) of their AI system, and they can get an objective view of the safety and desirability of their AI solution (Fig. 3).

- **Privacy score (P):** How well the AI system performs in protecting user privacy?
- **Ethics Score (E):** How compliant (or trained) the AI system is in preserving human values of dignity, fairness, respect, compassion and kindness for a fellow human being. Does the system have a preferential sense of duty towards children and vulnerable people like elderly, pregnant women and sick. How well does it value environmental sustainability, green energy and sustainable living?
- **Diversity Score (D):** How well the system is trained for diversity in race, gender, religion, language, colour, features, food habits, accent etc.?
- **Equity & Fairness Score (E):** Does the system promote equity and treats everyone fairly?

- **Auditability & Transparency Score (A):** How good is auditability of decisions made by the autonomous system? Can the decisions taken be explained?
- **Consistency across geographies & societies score (X):** How good is the AI system in delivering expected results across geographies and across different societies? Does it work for the low resource communities? Does it work across the Digital divide?
- **Misuse Protection Score (M):** Has the system been designed to incorporate features that inhibit or discourage the possible misuse? Are the misuse protection safeguards built into the system?

### 3.5.3.7. An integrated view of DEEP-MAX scorecard with TAM-DEF framework for AI Why DEEP-MAX scorecard parameters are slightly different from TAM-DEF framework components

The DEEP-MAX scorecard has been deliberately chosen to be slightly different from TAM-DEF framework. It could be noticed that one of the six components of the TAM-DEF framework, namely Accountability and Legal issues, has been kept out of DEEP-MAX scoring system. One can also observe that two of the TAM-DEF framework components have been split into two scores per component. Fairness and Equity component of TAM-DEF framework is split into two scores of Diversity and Equity. Similarly, Ethics component has been also split in two scores of Privacy and Ethics (Fig. 4).

This has been done with twin objectives of:

- Making the critical concerns regarding AI systems as an explicit Scorecard element. For example Diversity training of AI modules is absolutely must before any AI system is allowed to interact with people or to make decision about people. Therefore it has been treated as a separate element for scoring purpose – carved out from Fairness and equity component. Similarly, privacy in the age of data is possibly the single most important concern under Ethics component, hence this also has been treated as a separate element for scoring purpose, again carved out of the Ethics component.
- Keeping the Scorecard practical and computable. While clear understanding of Accountability and Legal issues are important for public use of AI systems, this component lies mostly in the domain of Law and does not lend itself easily towards scorecard mathematics. This issue gains prominence in those autonomous AI systems

DEEP-MAX Scorecard for AI under TAM-DEF framework		
<b>D</b>	Diversity	Diversity score – how well the AI system is trained for diversity in race, gender, religion, language, color, features, food habits, accent etc.?
<b>E</b>	Equity & Fairness	Equity score - Does the system promotes equity and treats everyone fairly?
<b>E</b>	Ethics	Ethics score – how well the AI system preserves human values of dignity, fairness, respect, compassion and kindness for a fellow human being
<b>P</b>	Privacy & Data Protection	Privacy & Data Protection Score - how well the AI system protects privacy of individuals? Does it have data protection features built in?
<b>M</b>	Misuse-protection	Misuse prevention score - Has the system been designed to incorporate features that inhibit or discourage the possible misuse?
<b>A</b>	Audit & Transparency	Auditability score - How good is auditability of decisions made by the autonomous system?
<b>X</b>	Cross Geography & Society	Cross Geography & Cross Society score – How well the AI system works across geographies and across societies especially for the disadvantaged societies?

Fig. 3. DEEP-MAX scorecard for AI under TAM-DEF framework.

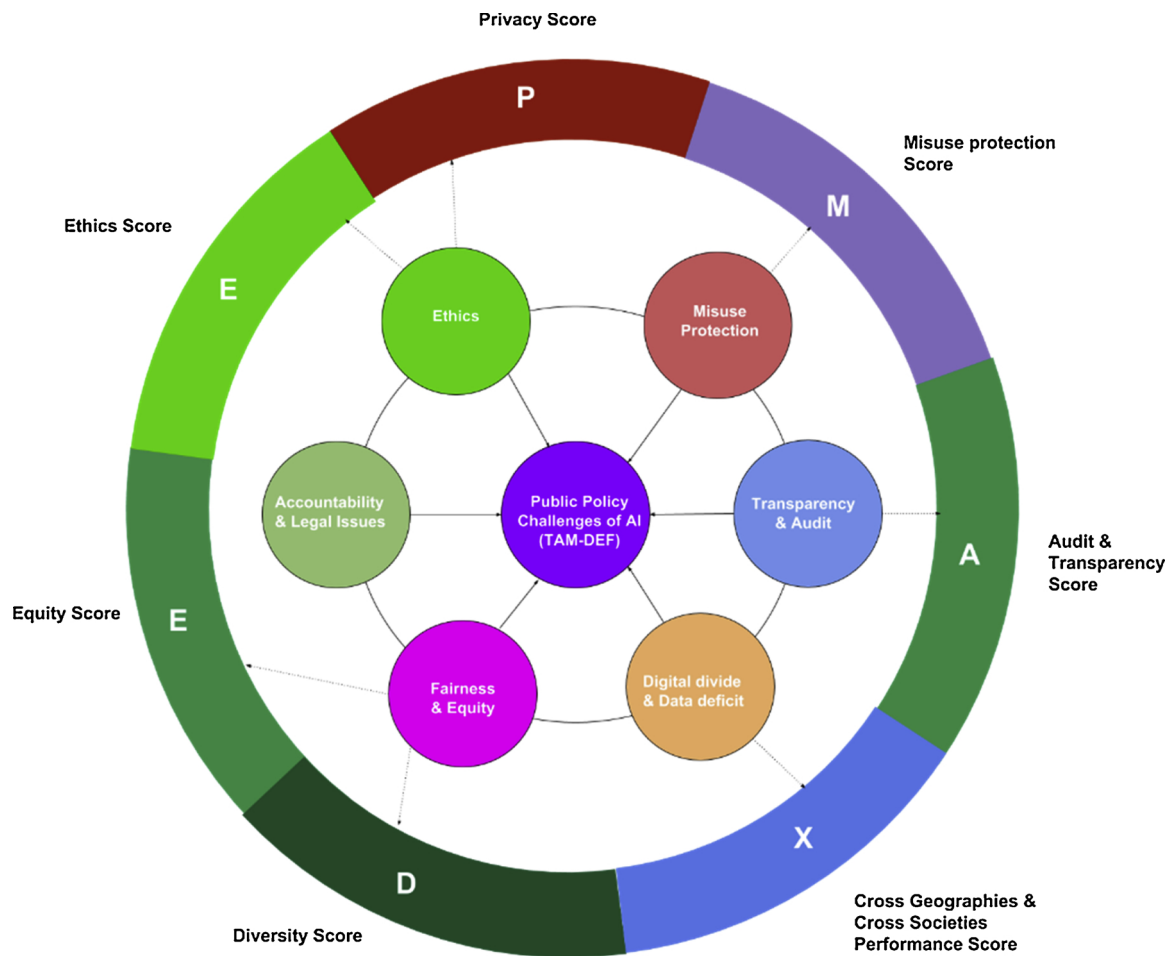


Fig. 4. An integrated view of DEEP-MAX Scorecard under TAM-DEF framework for Public Policy challenges of AI.

where irreversible decisions are allowed to be taken. The specific Laws of the countries start governing such usage. Given the variance in Laws and differences across the nations, this component of TAM-DEF framework has been kept out of DEEP-MAX scorecard.

#### Blockchain for safe and TAM-DEF compliant AI

- **Training data certification:** Blockchain can provide a trusted mechanism to certify the quality of training data for an AI system module. Whether a given AI system or module has been trained using a diverse data set incorporating race, gender, language, ethnicity, religion, and other forms of diversity or not can be easily verified if the certification is done using Blockchain (AI Certification Transparency & Scorecard Blockchain – ACTS Blockchain)
- **Tamperproof DEEP-MAX scores:** Any AI models developed, should be tested on a set of standardised data sets, each measuring one of the 7 DEEP-MAX Scores discussed above. These scores would be put on the ACTS – Blockchain and each AI module when shipped would ship with this trusted score card along with its training data certificate.
- **Activation Atlas based AI rating system:** One of the key areas of current research in AI is understanding decision making inside the neural network (Carter et al., 2019). The activation atlas of an AI model correlates the internal neural net nodes into features, and a visual overlay of the features help improve our understanding of the AI decision making process. The ACTS-Blockchain would carry the activation atlas information of the AI system along with its certification and DEEP-MAX score. This activation atlas information can be used for alerting users of the potential pit falls of the AI model.

The activation atlas would help explain the decision making process of the AI module and thereby add to the transparency. It also can help explain the cause for poor DEEP-MAX scores.

- **Built in Misuse prevention using Blockchain:** For public policy practitioners, misuse protection of AI systems is possibly the biggest challenge. A face recognition AI system for apprehending dangerous criminals can be easily tweaked for unscrupulous use, especially in less developed democracies. In this case a Blockchain based record keeping for any substitution or changes in the criminal image database would help safeguard the system from possible misuse. The Blockchain would contain a tamperproof record of the changes made along with the authorisation details, making all changes traceable.

**3.5.3.8. Periodic update of DEEP-MAX Scores of AI modules in public use.** One of the unusual challenges of the AI systems is that many of them keep learning as they are being used (after initial training). In such cases there is a strong likelihood of their behaviour changing as they process more data. It, therefore, becomes imperative to do a periodic testing of such AI systems and updating their scores on the DEEP-MAX scorecard. The periodicity of update would be dependent upon the class of use case and degree of autonomy granted to the AI system.

This DEEP-MAX rating system under the TAM-DEF framework is key to safety and desirability of AI systems for public use. This scorecard is critical because most AI programmes are likely to be used as off-the-shelf components for building a more complex AI systems. If a poorly designed AI component, which scores low on say diversity, is used in a more complex system say crime prevention, the results can be

devastating. It can result in racial or gender profiling, denial of access to financial institutions to persons residing in a particular pin code or locality, unfairly charge higher insurance premium based on a personal characteristic etc.

### 3.5.3.9. Research agenda

- The above discussion leads to a rich area for further research by scholars. Each of the three strategies listed above for tackling the public policy challenges of AI is a fertile ground for further research.
- *AI Standards and Rating*: A Global alliance to democratically and transparently standardise and rate AI applications is urgently needed. What should be its structure, how the rating standards should be chosen, how to account for global diversity and cultural norms, how to ensure that standards are followed – these all are significant questions to be answered in future research.
- *Designing Data Sets for each of the 7 DEEP-MAX scorecard tests*: AI systems today span a wide variety of applications like computer vision, autonomous navigation, medical intervention, text analysis, speech analysis, financial decision making, and education and testing. Even though they all work on the same underlying fundamental, they each need to be trained on a very different class of datasets, some need images, some voices and others just numbers. It is a significant work to design benchmarking databases which would generate a reliable and transparent DEEP-MAX scores for a given class of AI applications.
- *Integrating Blockchain for Trust and Safety of AI*: This the second area of further research. How do we make AI modules trusted and well understood worldwide. How do we ensure their DEEP-MAX scores are not tampered with. Whether verifiable attempts have been made during training of the AI modules to comply with each of the 6 checkpoints of TAM-DEF framework. Designing the ACTS-Blockchain for AI applications is another critical area for further research.
- *Design AI for protection against Misuse*: As discussed above, one of the ways to prevent misuse of AI is to make the misuse prevention as a built in feature in the design phase itself. As suggested above one can make new data additions or deletions (for example – faces to be identified in a crime prevention system) a permanent record on a Blockchain backbone with clearly identifiable individuals who ordered the change along with date and time stamp.

### 3.5.4. Governance of AI and connected systems – Marijn Janssen

During the last decades, information systems have become increasingly interconnected. What started with the Internet has evolved into the Internet of Things (IoT), where sensors and actuators are interconnected to measure and control systems from tooth-brushes to complete factories and refineries (Lee & Lee, 2015). This goes along with the availability of more and more Big and Open Linked Data (BOLD) about temperature, traffic jams, geolocation, pollution, gas and water flows, force, acceleration, and production throughput (Janssen, Matheus, & Zuiderwijk, 2015). The data deluge has resulted in the adding of intelligence in the forms of algorithms to deal with these large amount and variety of (big) data.

AI has become an integral part of these connected systems, like autonomous cars, smart living environments, and smart energy applications for the energy transition. Within these systems, AI can be used for simple tasks like cleansing data, to complex decision making processes involving data from countless distributed sensors. The intelligence provided by the systems enable better information sharing and cooperation resulting in improved user-experiences and personalisation, higher levels of efficiency and a reduction of costs. The algorithms for creating intelligence are also scattered in the systems, they might be at the sensors to ensure that privacy-sensitive data is not shared, or to ensure fast reaction time (e.g. edge computing), or the intelligence might be in the data centres of companies. Edge computing

complements data processing by providing large number of distributed nodes close to the data source and end users (Morabito, Cozzolino, Ding, Beijar, & Ott, 2018). Often the algorithms for creating intelligence might be executed on the cloud owned by other players and all kinds of software can be used. This all results in an interconnected socio-material systems which integrate data, algorithms, people, processes and software (Janssen & Kuk, 2016b). The paradox is that AI systems become increasingly omnipresent, however at the same time become less visible. AI performs tasks and make decisions without that people are being aware of this. Within cars, smart phones and energy networks all kinds of AI is already used nowadays.

**3.5.4.1. Challenges.** Technology need to be governed to ensure that the benefits are gained and the risks mitigated. With new technologies determining which responsibilities are needed for ensuring proper functioning and development is often difficult, however, the more needed the more powerful a technology is to ensure clear accountabilities and to deal with the risks. Unclear dependencies between data and algorithms, shared roles and joint operation among departments and organisations strengthen the dilution of responsibilities. Who is responsible for proper functioning and avoiding the making of mistakes becomes unclear.

*Complexity, uncertainty and materiality*: Data is collected and stored at multiple places in different ways. Data is collected by different organisations using all kinds of sensors and transformed when processed (Janssen, Van de Voort, & Wahyudi, 2017). Often it is unclear for what purpose the data is collected, what the limitations of its use are, who the owner of the data is, if data owners have given consent for use, what the quality of the data is. A challenge is to understand the data provenance and to ensure good data governance in a complex network operated by many players who all have a piece of the puzzle.

*Ensuring the making of correct decisions*: Different areas of our daily activities are being digitally recorded and a variety of algorithms are used to process the data. Data is collected for the purpose to be used in decision-making. Data is often not collected using an experiment set-up or another way of systematic research and the (lack of) availability of data influence the outcomes. Algorithms are not designed to deal with the dynamics and variety of inputs and might result in wrong outcomes. Data can be leading and the data bias can result in the inability to replicate studies, compromise the generalisability (Janssen & Kuk, 2016a) and result into wrong decisions.

*Who is responsible?* As more and more technology is interconnected, it is hard to establish a causal relationships between an event and a failure. For instance, who is responsible if the algorithm provides incorrect outcomes due to some anomalies in data that is collected by multiple sensors? Sufficient data quality is a condition for using the algorithm, however, perfect data quality probably does not exist. The data providers can make the argument that data quality is never 100% and even have included this contractually, whereas the algorithm provider can blame the data. Another example is the question about responsibility for a decision made by a deep learning application in which the causality of how deep learning applications arrive at the decision is now known. Such questions raise further questions of what the responsibilities of the designers are and what the responsibility of the users are? What can we expect from computing technology? Should societal values be included in the design of AI? Where should we add for the checks and controls to prevent the making of mistakes and to ensure that mistakes are detected?

*Lack of governance*: Besides its obvious advantages, AI holds risks for society. Algorithms may develop biases due to measurements problems, their training data, reinforce historical discrimination, favour a political orientation, reinforce undesired practices or result into unanticipated outcomes due to hidden complexities (Janssen & Kuk, 2016b). Governance is needed to unravel the complexity and to understand how connected AI systems influence our decision-making.



**3.5.4.2. Opportunities.** Algorithms can be embedded in our daily life. Algorithms warning people to avoid collisions in cars and algorithms that assist with the efficient use of our washing machines when electricity prices are low. There is an abundance of enthusiasm and optimism about AI data can be used for good. The dual use of data makes it possible to advance our society, but also to suppress the poor. At the same time the emerging AI-based systems often lack transparency, accountability and oversight. A new area of data and AI governance is needed to ensure that the benefits can be gained and risks avoided. Value-aware AI systems need to be designed that ensure that decisions are made correctly, that societal values and norms are represented in AI systems and people can safely enjoy the benefits of AI.

**3.5.4.3. Research agenda.** AI results in connected algorithmic systems in which often a number of AI techniques are combined and multiple sources of data are used and computing can occur anywhere. These systems are used more and more to make critical decisions, but the decisions might not always be correct.

**A systems perspective to unravel complexity:** Stacking components on top of each other combined with connecting them at different layers creates a complexity in which cause-and-effect relations are hard to understand and predict. This undermines the governance and accountability. The approach to tackle AI from an architectural view needs to take a systems perspective for understanding and controlling the complexity.

**Dealing with uncertainty and various quality:** The environment will always be subject to changes and there will be uncertainty about its development. Furthermore information sources have various degrees of quality and might be collected for another purpose than it will be used for. Therefore AI should be able to make sound and robust decision in uncertain and complex environments in which information have various qualities. Connect systems should be designed in such a way that bias in data is avoided and reliable decisions are made.

**Value aware systems:** Society norms and values should be represented in the AI systems. The EU General Data Protection Regulation (GDPR) is a regulation that applies to AI, as it states that automated algorithmic decision making should be explainable to persons who are affected by it. This is an important value for European citizens. Although norms and values differ per society, there are universal values that can be adhered to and embedded in the systems.

**Compliance-by-design:** The best way of ensuring embedding of values and regulation in the AI systems is to ensure that these are taken into account from the very start of the design process and that the system ensures that the values and regulations are adhered to. For example, the public is warned when facial-recognition systems are being used to track them, and that they should have the right to reject the use of such technology.

**AI governance:** People and organisations design and operate connected AI systems. AI governance should ensure that the right value are embedded in the systems. Autonomous systems need to be governed, but also the network of interconnected systems need to be governed. AI systems are not designed to last forever as they evolve with the environments and data, algorithms, people, processes and software

influence each other. Governance is needed to deal with bias in the data introduced, information is missing, when data is stolen, AI systems are taken over by criminals and so on. Sound governance is needed in which clear responsibilities are defined and risks are assessed. Robust AI governance are needed to deal with the above challenges.

#### 4. Discussion and recommendations for future research

The expert views outlined in the previous section are grouped in alignment with a number of perspectives on AI: Technological; Business and management; Arts, humanities and law; Science and technology; Government/public sector. This section pulls together many of the key themes and significant factors arising from the individual contributions to develop an informed discourse on many of the key topics and potential for future research.

##### 4.1. Challenges and opportunities

The individual perspectives have highlighted several challenges and potential opportunities relating to AI within a number of different themes and applications. [Tables 4 and 5](#) highlight each of these areas.

The increasing complexity of AI and the increasing number of genres of application where the technology can be applied is growing at pace. The disruption potential is vast, led by a momentum of change where many of the normative rules of governance and transparency need to be reconfigured to cater for the complexities and impact of AI. Visualising this complexity in terms of a transparent perspective of the underlying algorithmic architecture, particularly in the era of deep learning systems, is problematic ([Reza Tizhoosh & Pantanowitz, 2018](#)). [Edwards](#) highlights the significant challenges in the explainability of systems and algorithms that underpin AI technology and debates within the literature on transparency vs black box perspectives. The human trait of explaining the underlying reasoning behind a decision and applying this same logic and requirement in the context of AI is complex ([Miller, 2019](#)). [Edwards](#) highlights that the need for explainability in the age of AI is perhaps not a universal requirement and cites the criticality of accuracy over transparency within medical diagnosis, where a black box approach based on an evidence approach is deemed to be acceptable ([London, 2019](#)).

The perspectives from [Walton](#) question the readiness of organisations to make the transition to AI, highlighting limitations in exiting information processing and the importance of adaptiveness for successful transition. The perspectives highlighted in [Kar](#) introduce similar themes where the abilities of organisations in the context of people and process maturity are not yet mature enough to exploit the full potential of AI. Studies have highlighted that organisations face significant issues where the lack of an effective strategy for human vs AI interaction could affect critical business areas and fail to address concerns from the human workforce ([Fry, 2018](#); [Sun & Medaglia, 2019](#)). AI can help organisations to develop operational and strategic awareness but information quality is a critical component ([Westerman et al., 2014](#)) for effective change. [Kar](#) raises the prospect of a potential information and technological divide between large and smaller organisations perhaps

**Table 4**

Focus areas of the 16 funded Centres for Doctoral Training in the 2018 UKRI AI CDT funding call.

Focus area	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Healthcare, biomedical	✓	✓	✓	✓	✓											
Responsible AI, human						✓	✓	✓	✓							
Core AI research					✓					✓						
Fundamental science					✓											
Language processing											✓	✓				
Environment, sustainability													✓	✓		
Engineering, nano-devices															✓	
Creative industries, music																✓

**Table 5**  
AI challenges.

Title	AI Challenge	Contributor
Assuring explainability	Explainability in the context of articulating the reasoning behind a particular decision, classification or forecast can be complex for AI based systems (Miller, 2019). Even if an explanation is given that appears valid, how do we know if its accurate? How could can this be tested? Process need to be in place for people to be able to challenge an AI based decision to ensure transparency and accountability.	John S. Edwards
Ecosystem boundaries	The potential exists for ecosystem boundaries to exist between AI and humans (Fry, 2018). As AI tackles more complex topics the ability to exchange complex information between AI and humans will become ever more important. This highlights the challenge of how can organisations ensure people and AI can work together successfully?	Paul Walton
Decision making using AI	How can humans and AI be complementary in organisational decision making and work symbiotically to augment and enhance each other's capabilities and what would be the implications of using AI for future strategic business decisions?	Yanqing Duan, John Edwards, Yogesh Dwivedi
Migrating towards AI based automation	Increasing levels of automation have directly impacted workers in many ways. Organisations are likely to be faced with four major challenges: 1) how to select tasks for automation; 2) how to select the level of automation for each task; 3) how to manage the impact of AI-enabled automation on human performance; and 4) how to manage AI-enabled automation errors.	Crispin Coombs
Impacts on labour	Technological development has been astounding for the last 40 years, roles have changed and new jobs have been created, but the need for labour has not diminished. The perceived challenges in reassignment of jobs, re-skilling workers have been managed organically as technology change has advanced industry. In the new AI era, new roles will be created either in support of AI or in the design or assurance of AI technologies.	Spyros Samothrakis
AI trade offs	Challenges exist for estimating the trade-offs between differentiation and commoditisation of AI. The economic returns of technology are highest when it reaches a maturity of commoditisation. But AI systems are also expected to evolve themselves as they learn from the contextual and sticky knowledge within organisations, highlighting that perhaps AI can never truly be commoditised.	Arpan Kar
Digital marketing	Challenges exist for the adoption and implementation of AI for digital marketing. Areas such as: availability of data, required financial resources and trust where inherent fear on the prospect of AI and its adoption for marketing communications.	Emmanuel Mogaji
Implications for sales	As AI develops the mechanistic nature of algorithms employed in sales based systems may reduce the need for competitive differentiation, creativity and interaction in sales exchanges. The net effect of this could be reduced investment in sales training and development.	Kenneth Le Meunier-Fitzhugh & Leslie Caroline Le Meunier-FitzHugh
Impact within emerging markets	Within emerging markets the lack of education may be a significant challenge and a barrier to greater levels of AI adoption. The enterprise owners are likely to be less educated and the absorptive capacity and ability to understand the potential could be a drawback. If AI deployment is possible only in enterprises that are larger with a threshold amount of technological capabilities, the scale of adoption in future is likely to be low within these markets. The challenge within emerging markets will be for governments to embrace AI developments whilst being cognisant of the impact on replacing workers.	P. Vigneswara Ilavarasan
People centred perspectives	Challenges exist in the perception of AI in the context of adoption and implementation of the technology. To many people, AI is a concept that is hard to define and difficult to understand how it can manifests itself within in their everyday lives. Many people associate AI with negative press and media campaigns. This ill-defined concept and poor media coverage has resulted in negative association and poor brand image. The transition phase between things working with AI capability, and reduced 'smart' capabilities (augmented intelligence) can be confusing, frustrating and discriminating. Challenge remain on how 'natural' the interaction with AI can be.	Jak Spencer
Social and cultural aspects.	AI might be more efficient and reliable, but may face social resistance at least for some considerable time. if AI is accepted with enthusiastic fashion, it might explosively spread before the realistic implications from its use are known. How will the cultural distance between humans and AI impact people's demand for AI products? Are AI based products real substitutes for human derived products according to consumers' perceptions? Which products and services will be affected most from the lack of human proximity between the labour employed in these goods and service and the consumer?	Annie Tubadji
Perspectives from the fundamental sciences	The use of ML has exploded and it is now employed in most branches of fundamental science, with increasing success and acceptance. Due to the speed with which AI has evolved, it may be difficult to commit to a specific software framework and embed it in within existing analysis packages, before the field has moved on.	Gert Aarts and Biagio Lucini
Terminology	The "artificial" in AI is by itself a terminological challenge. Similarly, several cognitive scientists, cyberneticians, system theorists, AI/robotics specialists, and sociologists argue that AI is merely indefinable because we do not have any good understanding of the word intelligence. The AI hype and ill-definition diverts focus from problems which should be prioritised instead of policy discussions having to do with robotic liability, and so on which take non-experts' accounts as expert knowledge and science fiction perspectives. The lack of empirical data makes all speculation a challenge in its own right. The most fundamental challenge is that despite the fact that AI has been hyped in the last five years, a disciplined turn to specialists and the grounding of research agenda on the basis of technical evidence should be a core priority of any work dealing with the politics and economics of AI.	Vassilis Galanos
Algorithmic challenges in the public sector	A number of challenges related to AI adoption in the public sector are not unique to AI, but instead overlap with well-documented problems of adoption of any new emergent technology in government. These classic challenges include: the quest for data integration across different organisations, resistance to change and threats of worker replacement. Challenges exist in the areas of AI algorithmic bias and opacity where citizen expectations of transparency and accountability need to be taken account of in the personal and political context. Who is	Rony Medaglia

(continued on next page)

Table 5 (continued)

Title	AI Challenge	Contributor
SMEs and public sector	accountable if a decision has been outsourced to an AI application and what is the citizen recourse when wrong decisions are made? SMEs and public sector companies may face many challenges in leveraging AI enabled tools when compared to large technology organisations. There are many challenges in adopting AI based tools ranging from data quality, privacy/security risks and to the shortage of qualified AI experts	Sujeet Sharma and JB Singh
Public policy changes	The Public Policy is facing unprecedented uncertainty and challenges in this dynamic world of AI. Everyday a new application based on AI is invented and unleashed to society. The velocity, and scale of AI impact is so high that it rarely gives public policy practitioners sufficient time to respond. Public Policy, by definition, needs to put in place regulations against a possible future development which could be detrimental to human values. This creates an interesting tension between the need to predict AI impact and inability to draw boundaries around this highly dynamic technology.	Santosh K Misra
AI governance	Technology need to be governed to ensure that the benefits are gained and the risks mitigated. With new technologies determining which responsibilities are needed for ensuring proper functioning and development is often difficult, however, the more needed the more powerful a technology is to ensure clear accountabilities and to deal with the risks. Unclear dependencies between data and algorithms, shared roles and joint operation among departments and organisations strengthen the dilution of responsibilities. Who is responsible for proper functioning and avoiding the making of mistakes becomes unclear.	Marijn Janssen

less able to innovate via AI.

The change within society from humans to intelligent machines making key decisions on medical diagnosis, resource allocation and analytics based prediction amongst many others, is problematic. The challenges outlined by *Duan, Edwards and Dwivedi* assert the requirement to develop a more detailed and informed perspective on the implications and criticality of AI decision making on humans and to be cognisant on the cultural aspect. The need for a more informed debate on this topic is clear as we struggle to understand the impact of human vs machine interaction, the human enhancement capability and boundaries therein (*Miller, 2018*). The cultural perspective on AI decision making and the transparency of the underlying algorithms that support this are key for technology acceptance (*Gerbert, Reeves, Ransbotham, Kiron, & Spira, 2018*).

The many challenges and opportunities presented by AI are detailed in *Tables 5 and 6* respectively.

One of the frequently debates on greater levels of AI within industry and society, is the replacement of workers due to the increasing levels of automation (*Frey and Osborne 2017*). Whilst it is clear that lower skilled roles are likely to disappear, the literature is increasingly recognising that there is a need for humans in the loop (*Jonsson & Svensson, 2016*). The perspectives on job roles and labour hours from *Coombs* and *Samothrakis* respectively, argue that there is a continuing need for humans to work alongside AI technology and that research is required to effectively analyse what tasks to automate, over reliance on AI and failsafe capability in the event of AI failure. The perspectives from *Spencer* reiterate these points, stressing the need for humans to be at the centre of any AI development and the benefits of moving towards a more fair use of AI to enhance human lives. Researchers have supported the need for AI technologies to augment not replace the work of humans, to support key tasks and deliver greater levels of performance (*Davenport & Kirby, 2016*; Wang, Törngren, & Onori, 2015a; Wang, Li, & Leung, 2015b). Workers are likely to progress higher up the value chain to solve design and integration problems as part of an integrated AI and human centric workforce (*DIN & DKE, 2018*). A number of these points are analysed by *Ilavarasan* where the emerging market perspective is outlined in the context of challenges from AI within India. *Ilavarasan* posits that the lack of complementary asset availability acts as a barrier to AI adoption and the dichotomy of governments encouraging innovation whilst being cognisant of the labour displacement from AI technologies. The Indian government commitment to AI via the digital India initiative (*Niti Aayog, 2018*) is clear however, the balance between the social potential of AI vs the impact on workers is yet to be played out.

AI technologies have become an integral element of digital strategies with chatbots and intelligent predictive analytics now the mainstay of many of many organisations (*Juniper Research, 2019*). *Mogaji* discussed perspectives on AI and highlighted a number of factors that may hinder adoption within digital marketing. The key points from this perspective were the availability of data, financial resources and trust in AI where *Mogaji* posited these factors as significant challenges for the further use of AI within digital marketing. The recommendation in this perspective on the need for more focussed research on the integration of AI within organisations and the ethical considerations of the technology is supported within the literature (*Gupta & Kumari, 2017*; *Sun & Medaglia, 2019*). In a similar vein *Le Meunier-Fitzhugh* argues for a greater understanding of how AI is influencing B2B sales exchanges and the potential consequences of humans interacting with AI sales assistants. In an age where big data analytics integrated with AI can guide consumers through the sales process (*Juniper Research, 2018*; *Loring, 2018*), many questions remain on the ethics and implications of sales algorithms and the human vs AI interaction.

The ethics and transparency debate surrounding the introduction of AI is ongoing with studies analysing the implications of the technology within healthcare (*Houssami, Lee, Buist, & Tao, 2017*) governance and safety (*Zandi et al., 2019*). The perspectives from *Tubadji* posit the lack of scientific economic recognition on the potential social changes from the emergence of AI and Industry 4.0 (I4.0). The perspective asserts the importance of cultural proximity in the context of humans vs AI, where a greater emphasis on culture based analysis can provide insight to diffusion of AI technology within regions. The lack of interpretability of AI is highlighted in *Aarts and Lucini* where, from the angle of fundamental science, the perspective advocates the needs of unboxing AI algorithms in order to engender wider acceptance of the technology in wider contexts. Studies have highlighted the implications of lack of AI governance and the potential for unintended consequences (*Janssen & Kuk, 2016b*; *Zandi et al., 2019*). *Janssen* argues for the criticality of AI governance not just at the algorithm and system level but also across network of interconnected systems and data levels.

The universal adoption of innovative technology by governments and use within the public sector is problematic within the IS and political context (*Eggers et al., 2017*). The perspectives from *Medaglia* stress the challenges of AI adoption within government asserting the criticality of dialogue with citizens in countering distrust and social applications of the technology and the assessment of AI readiness. Studies have hypothesised on the readiness of AI systems to perform manual government functions such as bail hearings, asserting that the technology is robust enough to deliver performance benefits over

**Table 6**  
AI opportunities.

Title	AI opportunities	Contributor
Modelling explainability	In the fields of medical diagnosis and treatment, explainability is perhaps less important than accuracy. Opportunities exists in conceptualising AI in the context of a black box approach where outputs should be judged using clinical trials and evidence-based practice to strive for accuracy (London, 2019).	John S. Edwards
Organisation effectiveness	There are a number of opportunities for organisations to utilise AI within a number of categories: organisational environment, operations, Interaction, case management automation, governance and adaptiveness. AI can provide the opportunity for organisations to develop both operational and strategic situation awareness and to link that awareness through to action increasingly quickly, efficiently and effectively.	Paul Walton
Transformational potential of AI	Opportunities exist for the development of a greater understanding of the real impact of decision making within organisations using AI in the context of: key success factors, culture, performance, system design criteria.	Yanqing Duan, John Edwards, Yogesh Dwivedi
Automation complacency	Although, automation complacency and bias can speed up decision making when recommendations are correct. In instances where AI provides incorrect recommendations, omission errors can occur as humans are either out of the loop or less able to assure decisions. Opportunities exists to explore and understand the factors that influence over reliance on automation and how to counter identified errors.	Crispin Coombs
Workforce transition	Society is likely to be significantly impacted by the AI technological trajectory if as commentators suggest, society achieves full automation in the next 100 years (Müller & Bostrom, 2016; Walsh, 2018). The opportunity here for organisations and government, is the effective management of this transition to mitigate this potentially painful change.	Spyros Samothrakis
Enabler for platforms and ecosystems	The exploration of opportunities as to how AI can be leveraged not only at the firm level but as an enabler in platforms and ecosystems. AI may help to connect multiple firms and help in automating and managing information flows across multiple organisations in such platforms. Significant opportunities exist for AI to be used in such platforms to impact platform, firm and ecosystem productivity.	Arpan Kar
Enhanced digital marketing	AI offers opportunities to enhances campaign creation, planning, targeting, planning, and evaluation. AI offers the opportunity to process big datasets faster and more efficiently. Opportunities exist for more innovative and relevant content creation and sharing using AI tools and technologies.	Emmanuel Mogaji
Sales performance	Opportunities exist for improving the sales performance using AI driven dashboard, predictive and forecasting capability and use of big data to retain and develop new customer leads. Additionally the use of AI algorithms can contributing to productivity and provide sales process enhancement through elimination of non-productive activities and removal of mundane jobs.	Kenneth Le Meunier-Fitzhugh & Leslie Caroline Le Meunier-FitzHugh
Emerging markets	The presence of complementary assets are likely to influence the transition to AI in the developing world. Opportunities exist for the lessons learnt from India and Kenya to benefit similar low income countries in future. For instance, Pakistan, Vietnam, and others are imitating the success story of the Indian software services exports story.	P. Vigneswara Ilavarasan
People centred AI	AI can potentially be used to enhance 'softer' goals rather than the drive to economic productivity or efficiency. The genuine needs of people can be identified that can solve real-world problems. As our interactions with machines start to become more and more human-like, the opportunity lies in the design of new personalities and the creation of new types of relationship.	Jak Spencer
Taste fear and cultural proximity	Opportunities exist in the focus on market taste, fear and cultural proximity to improve organisational use of AI. While their attention is currently focused on the pros from efficiency gains, they might be overlooking the market reaction to the integration of AI in their production process. Learning about tastes informs the market about AI-generated products and services. Learning about fear within AI-related social opinions and policy-making tendencies can help us make evidence-based AI-related decisions. Learning about the importance of cultural proximity in the context of AI-human cultural distance can help to quantify the cultural gravity effect that bounds our consumption of AI-goods and products.	Annie Tubadji
Power of AI algorithms	ML can be broadly understood as an optimisation problem, in which the parameters of a model function are selected to reproduce as closely as possible a known response. This problem requires the use of computational resources. The availability of algorithms that are as fast as possible becomes paramount. Computationally power-hungry problems of this type have been well known in Science and Engineering, via the use of parallel programming and use of a Supercomputer. This approach using high-performance computing (HPC), is offering the possibility to accelerate ML algorithms by orders of magnitude, to a point when a prediction can find timely applicability.	Gert Aarts and Biagio Lucini
Accurate narrative	Opportunities exist to impose realistic expectations of AI. Far-fetched expectations have been harmful and contributed to the confusing narrative on AI. The history of AI shows a repetitive rise and fall pattern of hype and disillusionment; large availability of grants followed by long periods of research support stagnation – this happened due to AI specialists, in their attempts at establishing their field have made very brave and overly ambitious (and ambiguous) promises to eventually remain unfulfilled. Given the evidence of the current negative effects of non-specialist intrusion, the right to intense boundary work to separate who is entitled to be a spokesperson of AI and who is not, should be made.	Vassilis Galanos
Fostering citizen trust	AI applications, such as rules-based systems, speech recognition, machine translation, computer vision, machine learning, robotics, and natural language processing, have the potential to free up precious cognitive resources of public workers, which can then be allocated to tasks of higher added value. Opportunities exist for AI applications to foster citizen trust. Unfair, inefficient, or even distorted provision of government services can be potentially reduced by the use of AI.	Rony Medaglia
SMEs and public sector	The potential of AI within SMEs is in the automation of various tasks with decision making components such as in the functions of finance and customer services. AI applications could	Sujeet Sharma and JB Singh

(continued on next page)



Table 6 (continued)

Title	AI opportunities	Contributor
	help SMEs in matching customer invoices with received payments, AI chatbots could help enterprises answering customer's simple requests. AI could improve the efficiency of SMEs, automating a number of business processes Public sector and government organisations generate lots of data through their processes and hence more potential exists for the implementation application of AI technologies. AI could help in identifying the target citizens for welfare schemes and payments. The judiciary in developing countries could be improved where cases are pending from decades due to availability of limited resources. AI could help in deciding on the bail hearings in courts as machine learning technologies are now robust for such applications.	
Public sector benefits	Opportunities exist in governments throughout the world via the use of AI to tackle problems such as: shortage of resources, scale of operations and standardisation of government delivery systems. Governments could transform their service offerings by using AI to address all the these issues. Governments can offer benefits to citizens via the use of: smart services, intelligent adaptive forms and predictive service delivery.	Santosh K Misra
AI governance	Algorithms can be embedded in our daily life. Algorithms warning people to avoid collisions in cars and that helps us of use our washing machine when electricity prices are low are much desirable. There is an abundance of enthusiasm and optimism about AI data can be used for the good. The dual use of data makes it possible to advance our society, but also to suppress the poor. At the same time the emerging AI-based systems often lack transparency, accountability and oversight. A new area of data and AI governance is needed to ensure that the benefits can be gained and risks avoided. Value-aware AI systems need to be designed that ensure that decisions are made correctly, that societal values and norms are represented in AI systems and people can safely enjoy the benefits of AI.	Marijn Janssen

existing processes (Mullainathan & Spiess, 2017). Misra asserts that due to the large number of transactions and regulatory services, governments are likely to be one of the largest adopters of AI. The perspective highlights the lack of an AI comprehensive public policy framework and presents the TAM-DEF framework to objectively test AI validity prior to procurement.

The individual perspectives from the invited experts and wider literature have offered unique insight to the subject of AI from a number

of viewpoints. Each of the contributions offer a number of potential research opportunities based on an assessment of research agenda in the context of each perspective. Many open questions remain on a number of aspects of AI:

1. The literature seems to conclude that the future of AI requires humans in the loop and that AI should be seen as augmenting the potential of humans not replacing them. However, is the concept of

Table 7

UN sustainable development goals vs AI technology driven change.

UN sustainability goals	AI technology potential in delivering UN goals.
No poverty Zero hunger Good health and well-being	The implementation of AI technology is likely to drive increasing levels of automation within manufacturing with resulting impacts on emerging and developed economies. Studies have highlighted the inevitable loss of low skilled labour and potential creation of new higher value jobs where human cognitive related skills can be utilised within the workplace. This is predicted to disproportionately affect many of the emerging Asian economies that have traditionally relied on this type of work. However, as new roles are created to support the increasing use of AI, requiring new skills and training, this realignment is likely to have a beneficial impact on raising peoples quality and standard of life. Within many emerging economies, particularly in rural areas, medical practitioners are in short supply. AI based diagnosis systems could be utilised to support doctors and potentially speed up the treatment process leading to health benefits for the population.
Quality education Gender equality Reduced inequalities	Education is likely to be impacted by the emergence of AI. Schools and universities could utilise AI technology in the classroom to aid the learning process and assist educators in their interaction with students. The Japan based study by Hamaguchi and Kondo (2018) highlighted the disproportionate impact on female workers from technology adoption when compared to male workers. These impacts could potentially worsen within the AI era unless positive steps are taken by policy makers and governmental organisations. Greater faith in AI systems could reduce inequalities due to the inability of potential bribery, intimidation and transparency as long as algorithms are open and certified.
Clean water and sanitation Affordable and clean energy	AI technology has the potential to predict energy and utility demand and react to climate change using big data and intelligent energy supply systems. The net effect of this change would be less waste, a more efficient supply network and lower cost energy, water and a means of assuring and promoting economic development amongst the world population (Cohen & Kharas, 2018).
Decent work and economic growth Industry innovation and infrastructure Sustainable cities and communities	Work, economic advancement and the growth of industry will be impacted by the adoption of AI technologies. Greater levels of automation and the advancement of machine learning technologies will improve working practices and productivity. This will in turn drive increased worker skill levels and growth within a number of sectors. The use of AI can engender innovation and greater levels of sustainability as governing authorities strive to incorporate AI technologies within communities and cities.
Responsible consumption and production Climate action Life below water Life on land	Responding to climate change and resulting impact is often costly. The poor are all too often the first to be impacted by climate change and, for the most part, will suffer the most in terms of loss of welfare and opportunity. AI technology improves the quality of understanding and responding to climate impacts and could end up being a vitally important part of assuring and promoting economic development amongst the world's least well off (Brookings Institute, 2019). The potential improvements to forecasting and modelling via the use of machine learning elements of AI and big data, can directly contribute to the ongoing human impact on use of valuable resources, life below water and on land. This use of technology can potentially force human change in these areas as AI systems help to gain consensus on key global sustainable issues relating to the United Nations Framework Convention on Climate Change (UNFCCC), subsequent Kyoto Protocol (2013–2020) and Paris Agreement (2015).
Peace justice and strong institutions	The combination of AI technology and human in the loop capability could potentially reinforce peoples trust in areas such as: medical diagnosis, interpretation of law and statute as well as government institutions that can be made more effective and efficient via AI technology.
Partnerships for the goals	Partnership between institutions and decision makers is required at an international level to enable acceptance of AI and for the technology to deliver the required development outcomes.

workers moving up the value chain to higher skilled jobs a universal one, especially within emerging economies?

2. Governance of AI technology is a key prerequisite prior to widespread adoption within industry and government. It is acknowledged this is likely to be a trade-off between transparency and performance. The Collingridge dilemma highlights this issue succinctly in the sense that by initiating greater AI oversight early in the AI lifecycle, could be relatively straight forward as the technology is relatively young and still hiding many of the unintended consequences. However, authorities could choose to wait until AI is relatively mature but then run the risk of losing control over its regulation (Collingridge, 1980). The rigorous audit of AI algorithms is likely to be complex and time consuming. How will these tasks be undertaken and is there a potential scenario where AI systems are tasked with auditing and testing other AI technologies?
3. In a scenario where many of the current computational constraints are overcome, the potential disruptive change from AI could be significant as industry and services migrate to a more automated machine based position. What are the cultural and societal implications of this change? What are the risks for the change in interaction and how will this impact the future of human decision making? Will the onset of AI impact how we approach education, training and skills acquisition?
4. The trajectory towards greater levels of automation is likely to benefit performance and productivity, but how are AI systems able to navigate the complicated human attributes of uncertainty within out of the box scenarios?
5. The speed of AI technology adoption is staggering and the ethical elements have yet to be fully contemplated and formalised. What ethical protocols need to be designed and agreed as a matter of urgency and what ethics controls need to be developed along a roadmap of additional controls as AI expands further?
6. How can we ensure that humans are at the centre of AI design and development and that the future aligns with a more fair and equitable use of the technology to improve people's lives.

#### 4.2. UN sustainability goals and AI

The United Nations (UN) developed Sustainability Development Goals (SDGs) were developed in 2015 for the UNs vision for the future. The goals were presented as a blueprint and shared agenda for peace and prosperity for the planet and population. Seventeen SDGs have been developed to highlight many of the key themes relating to ending poverty, improve health and education, focussing on climate change, reducing inequality and developing sustainable economic growth (UN 2019). The study by Ismagilova et al. (2019), presented the UN SDGs in the context of future impact of Smart Cities and its citizens. The Hughes et al. (2019) study incorporated the UN SDGs from the perspective of blockchain technology and how this emerging technology could be aligned with the creation of business and social value (Hughes et al., 2019). This study has reviewed each of the UN SDGs from the perspective of potential alignment with AI and the major themes from this study. Based on these key comparisons, Table 7 details each of the SDGs and how AI technology can potentially align with each of the goals and deliver benefits as well as sustainability.

The alignment of the UN SDGs and AI technology and highlights the key factors that could benefit sustainability on widespread adoption. This is likely to require significant investment from governments and industry together with collaboration at an international level to effect governance, standards and security. The increasing use of AI has the potential to benefit many aspects of society in the longer term as humans are free to concentrate on tasks requiring greater cognitive load whilst more mundane jobs are performed by machines. However, although this vision of a society that benefits from the onset of AI is realistic, the short to medium term transition may negatively impact many vulnerable aspects of society. Governments and organisations

need to develop pragmatic strategies to educate and re-skill workers to ensure humans are not disenfranchised by the onset of AI within the workplace. The likelihood of humans remaining in the loop in conjunction with intelligent machines means that workers will still have a vital role to play within organisations, as AI based machines support human endeavours. The implementation of AI could benefit many of the UN SDGs directly and indirectly over time as society in general is changed to incorporate AI technologies. Society will be able to utilise AI technology to more effectively predict the impact that humans have on the environment and the planet. AI based systems will be used to establish fairness within institutions and remove the subjectivity and corruption that can be a barrier to citizen and government interaction within many countries.

#### 4.3. Future research agenda

Extensive opportunities exist for academic research within a wide range of topics pursuant to AI technologies and related impacts of the ongoing transition to use of intelligent machines within industry and society. Any future research agenda covering AI will be diverse in that the adoption of the technology impacts many facets of government and industry with wide implications for how humans will potentially live and work in the future. The research agenda detailed within Table 8 outlines potential areas of future research topics as outlined in the AI workshop held on 13th June 2019 at the School of Management, Swansea University UK.

The potential research agenda for AI technology in all its forms is significant. AI technology has tended to become a somewhat broad church where many forms of automation and limited intelligent machines are labelled as AI. The literature has sought to highlight a differentiation between task-specific, domain based AI and the more cognitive human centric form able to perform numerous intelligent tasks. The term increasingly used to describe this is: Artificial General Intelligence (AGI) – so called – real AI (Bostrom, 2011). The real future agenda and potential change within industry and society, is perhaps split along the lines of AGI and the more domain specific AI where key specific tasks will be performed by machines. However, current levels of technological advancement have yet to reach levels of what could be described as AGI and are not likely to reach this in the near future. The inherent complexity of the human brain has yet to be fully simulated by even the most sophisticated computer algorithms. Current research focus is predominantly focussed on domain specific AI as well as its potential impact on government, industry and society in general.

The cultural and societal impacts of further transition to AI technologies cannot be underestimated as people come to terms with machines taking on more tasks traditionally carried out by humans. The disruptive impact on many aspects of society including: manufacturing, logistics, education, interaction with government and health are all likely to affect workers in all these sectors. The potential benefits of AI systems may not be realised by all sections of society as a natural reticence to interact with new technology and perhaps fear of change, may limit transition in the short to medium term (Bostrom, 2011). Governments generally seem unable to keep pace in a regulatory context with the speed of AI innovation. Researchers have a valuable role to play here in the analysis of the many barriers to AI interaction and the psychological aspects of change in the workforce and society in general. Furthermore, the global impact of AI on emerging economies needs to be assessed via academic study to ascertain the likely impact on low skilled workers and the wider economies from greater levels of automation and machine learning systems. The regulatory issues also extend to algorithm assurance, governance and ownership of unforeseen outcomes as a consequence of poor algorithm performance and complexities (Janssen & Kuk, 2016b). How do we know what levels of testing and applied scenarios have been used to validate an AI algorithm? Are the key logic and execution paths transparent to decision makers to ensure they are comfortable with the performance and likely

**Table 8**  
Future research agenda for AI.

Policy and economy	
Title	Research agenda description
Impact on society industry and education	<ul style="list-style-type: none"> <li>• Many of the current research debates seem to be technological in nature and performance driven. A wider debate is required to take into account the cultural and societal impacts of AI technology and what it means in the context of peoples lives.</li> <li>• The literature has analysed the potential impact on many aspects of industry and citizen interaction. Here the advantages of AI are often presented as benefits to performance and productivity etc. However, further research is needed to ascertain how these benefits can be spread throughout society as a whole.</li> <li>• The potential impact that AI could have on education is a significant and consequential step that requires thorough analysis, detailed planning as well as effective assurance. If the student or pupil interaction with the AI system is processed with little or no human teacher governance, how can society as a whole be assured that learning is effective and not subject to inherent algorithm errors?</li> <li>• The impact on workers from further levels of automation and AI based technology has been widely commented on within the media and academic study. The levels of adaptation from within the existing workforce is as yet unknown. However, history tells us that as industry and society changes, humans generally adapt to the new ways of working and learning of new skills. Studies have articulated the extensive reach of many forms of AI within the medical and legal professions and well as manufacturing. Further research is needed to fully quantify the potential impact and how these functions will be performed either in a fully automated context or with humans in the loop within a creative destruction of jobs context.</li> </ul>
Regulatory implications for AI	<ul style="list-style-type: none"> <li>• The pace of change relating to AI technology is staggering. Although various departments may publish various technology related strategies periodically, government leaders generally do not seem to be cognisant of many of the key issues and implications for society and citizens. Furthermore, leaders seem to be slow to react to technological change demonstrating evidence of a knowledge gap and requirement for a cultural shift within the public sector.</li> <li>• The traditional policy of long term strategies from central government and public sector departments, does not work for fast changing technologies such as AI. Governments are better served by adopting a strategy for short to medium plans that can be flexible enough to cater for technological change and likely breakthroughs.</li> <li>• Little evidence exists that demonstrate governments possess any tangible strategy or depth of understanding to even begin to think on regulation of AI. The sanctioning of AI technologies within industry and government systems may be subject to different regulatory approaches depending on the perspectives of emerging vs developed markets. Institutions may delay the onset of AI technology if its implementation results in widespread job loss and disruption of societal norms. Academic study has a role to play here in the deeper analysis of the implications if AI systems and regulatory options within a global context.</li> <li>• The regulation of AI needs to factor in the problem ownership when things go wrong. Is it valid for deployment to never take lace unless an AI based system can be fully described and translated and how can this be fully assured?</li> </ul>
<b>Boundaries and awareness</b>	
<b>Title</b>	<b>Research agenda description</b>
Bias within AI	<ul style="list-style-type: none"> <li>• As human developers have written the algorithms that are used within AI based systems, it should not be a surprise that a number of inherent biases have slipped through into decision making systems. The implication for bias within AI systems is significant as people may end up being disenfranchised by incorrect logic and decision making.</li> <li>• What levels of algorithmic assurance are needed? How can humans trust a black box approach to AI? What levels of recourse to humans have if decisions are questioned? The further research in a number of these areas is critical as AI based systems become ever more complex and problematic to fully validate.</li> </ul>
Boundaries between AI and people	<ul style="list-style-type: none"> <li>• The societal impact of AI must not be underestimated especially as we reflect on the reality that only 50% of the global population has no digital footprint. What are the implications of greater levels of automation where workers operate using AI enhanced machines or interact with AI systems in the factory setting?</li> <li>• Although the safety aspect of people working in close proximity with machines is addressed in the workplace, the interaction element between workers and AI has not been addressed. This area needs further research with regard to the psychological implications and the medium to long term effect on humans required to regularly work closely with AI systems.</li> </ul>
<b>Making decisions with AI</b>	
<b>Title</b>	<b>Research agenda description</b>
Scientific problems towards achieving full scale AGI	<ul style="list-style-type: none"> <li>• The much hyped scenario of super intelligent forms of technology able to perform many of the cognitive tasks of humans across domains has not materialised and is not likely to in the near future.</li> <li>• Although the availability of big data in conjunction with AI has enabled greater levels of AI performance specific to key domains, it is widely accepted that AGI is perhaps a potential long term prospect if at all possible. Researchers have scaled back on areas such as autonomous cars and general AI cognitive ability across domains. Research in these areas should focus on the opportunities and implications for human enhancement via the use of AI to deliver heightened levels of human performance and abilities.</li> </ul>

(continued on next page)

Table 8 (continued)

Policy and economy	
Title	Research agenda description
AI and strategic decision making	<ul style="list-style-type: none"> <li>• The implications for AI technology being integral to strategic decision making are complex with significant implications if poor decisions are made. What levels of assurance are in place if AI systems can make significant decisions autonomously? If AI systems require a human in the loop for final assurance for key strategic decisions, what are the implications and risk to organisations for AI decisions deemed to be less strategic?</li> <li>• In the absence of true AGI can we outsource any strategic decisions to AI without appropriate checks and balances to deliver the required levels of assurance?</li> </ul>
<b>Future impact</b>	
<b>Title</b>	<b>Research agenda description</b>
AI leaders of the world (FAMGI – BAT: Facebook, Amazon, Microsoft, IBM, Baidu, Alibaba and Tencent) – how can society and government push back from this imbalance?	<ul style="list-style-type: none"> <li>• The huge tech companies in the US and China control a significant market share of the innovation and momentum within the fields of automation and AI technology. To a certain extent the lack of understanding and therefore, regulation from government and wider society has left a void that has been somewhat exploited by the high tech firms, with regulatory institutions seemingly playing catchup on AI.</li> <li>• Is society disadvantaged by this level of control of the AI research agenda by a small number of organisations? Could society as a whole benefit from greater regulatory or government involvement earlier in the AI product lifecycle at an algorithmic assurance level? Further research is needed in this area to identify potential frameworks and protocols that can identify how government and society can engender greater transparency for AI design and implementation.</li> <li>• Researchers have debated this topic for some time within the technology focussed literature. Some studies have articulated a negative narrative on the greater adoption of AI technology, whereas other studies have concentrated on the positive benefits without fully identifying some of the drawbacks to society as a whole.</li> <li>• Generally, more recent aspects of the literature have posited a more realistic view on AI advances acknowledging that we are far from delivering anything near full AGI. In a societal and cultural context, researchers have an important role in identifying the potential implications for emerging nations where workers could potentially be deeply affected by the onset of AI.</li> </ul>
The debate on AI being a force for good or bad.	

outcomes of the AI system? Who gets the blame when things go wrong? Academic research is needed to answer these questions in order to develop a deeper analysis of the potential implications for all key stakeholders.

The potential for inherent bias within AI algorithms and implications of humans in loop working in close proximity to intelligent machines, poses significant challenges in the context of trust, human safety and ethical considerations. We should not assume that workers will be comfortable with the AI enhancing human capability concept and that resistance as well as lack of trust is likely to be the norm within the workplace (Gupta & Kumari, 2017; Sun & Medaglia, 2019). These complexities pose significant challenges as organisations utilise the power of AI combined with big data for strategic and potentially autonomous decision making. Academic research has a role to play here in the empirical study of workers attitudes to trust and the deeper implications of human and intelligent machine interaction. The ethical and moral dimensions are potentially extensive especially in the context of organisation decision making. Is there an underlying cultural dimension to the ethics of AI logic and subsequent outcomes? Is there a potential trade-off where one attribute of an AI decision so important that another would be sacrificed? Which attribute would be deemed to be less important in the context of strategic decision making and how is this choice made? What checks and balances are needed to be in place for management to have confidence in AI decisions and recommendations? These are important questions and key topics within a potential AI focussed research agenda.

The significant innovation from the big technology leaders has somewhat driven the technological agenda for AI to the extent that most of society seems to be in catch-up mode as each new step is made. Is this the correct progression path for society as a whole? Is there a better model or framework that could engender enhanced levels of trust and understanding? Can wider sectors of society assess their potential

fears of AI in the context of real tangible benefits? Academic research could play a greater role in assessing the impact of this current model and develop a wider debate on the societal perceptions of the technology and speed of innovation.

## 5. Conclusions

In alignment with an approach adopted from Dwivedi et al. (2015b), this study presents a consolidated yet multiple perspective on various aspects of AI from invited expert contributors from public sector, industry and academia. The collective insights stem from the workshop titled “Artificial Intelligence (AI): Emerging Challenges, Opportunities, and Agenda for Research and Practice” held on 13th June 2019 at the School of Management, Swansea University UK. Each of the individual perspectives has highlighted the opportunities, challenges and potential research agenda posed by the rapid emergence of AI. Each expert was invited to set out their individual contribution in largely unedited form, expressed directly as they were written by the authors. This approach creates an inherent unevenness in the logical flow but captures the distinctive orientations of the experts and their recommendations. The key findings and open research question have been outlined and aligned with the academic literature.

The trajectory towards increasing applications using AI has the potential to change many aspects of human lives and impacting society as a whole. The way forward is not clear and the potential roadmap is undefined. There are numerous benefits that could accrue from AI but there are also significant risks that swathes of society may be disenfranchised from the implementation of the technology. Decisions made within the next few years on the forward path for AI are likely to have an impact on all our lives and the lives of future generations.



## Acknowledgement

This submission was developed from a workshop on Artificial Intelligence (AI), which was held at the School of Management, Swansea University on 13<sup>th</sup> June 2019. We are very grateful to everyone who attended the workshop and contributed their perspectives during the workshop and as an input to this article. We are also truly appreciative to those who although not able to attend the workshop, provided their valuable perspectives for developing this work. We are also very grateful to our Senior PVC – Professor Hilary Lappin-Scott, the keynote speaker – Mr Lee Waters AM, Deputy Minister for Economy and Transport, National Assembly for Wales and the following panellists from industry and public sector organisations for enriching our understanding of this emerging area by providing their valuable perspectives that have informed the views presented in this article: Ms Sara El-Hanfy, Innovate UK; Mr Peter Chow, AI & Cloud Product Design Manufacturing & Inspection, Fujitsu UK; Ms Kimberley Littlemore, Director of eHealth Digital Media, UK; Mr Chris Reeves, Country Digitisation Director, Cisco UK & Ireland; Mr Adam Wedgbury, Team Leader for Cyber Security Innovation, Airbus; and Mr Toby White, CEO of Artimus, Cardiff, UK. We are also very grateful to our colleagues, Amy Jones and Julie Bromhead, for all their valuable support for organising the workshop. Finally, we are grateful to the Emerging Markets Research Centre (EMaRC), Swansea i-Lab (Innovation Lab), and Department of Business at the School of Management, Swansea University for their financial support in the organising of this workshop.

## References

- Abarca-Alvarez, F. J., Campos-Sanchez, F. S., & Reinoso-Bellido, R. (2018). Demographic and dwelling models by artificial intelligence: Urban renewal opportunities in Spanish coast. *International Journal of Sustainable Development and Planning*, 13(7), 941–953.
- Abbot, J., & Marohasy, J. (2013). The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia. *WIT Transactions on Ecology and the Environment*, 171, 287–297.
- Acemoglu, D., Autor, D., Dorn, D., Hanson, G., & Price, B. (2014). Return of the Solow Paradox? IT, productivity, and employment in US manufacturing. *American Economic Review*, 104(5), 394–399.
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160.
- Adam, T., O'Sullivan, A. C., & Russell, R. (2017). *Artificial intelligence and public policy*. Report of Mercatus Center, George Mason University.
- Agrawal, A., Gans, J., & Goldfarb, A. (2017). What to expect from artificial intelligence. *MIT Sloan Management Review*, 58(3), 22–27.
- Aguiar, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, 91(1), 34–49.
- Al-Emran, M., Mezhyuev, V., Kamaludin, A., & Shaalan, K. (2018). The impact of knowledge management processes on information systems: A systematic review. *International Journal of Information Management*, 43, 173–187.
- Anderson, K. P. (2019). Artificial intelligence-augmented ECG assessment: The promise and the challenge. *Journal of Cardiovascular Electrophysiology*. <https://doi.org/10.1111/jce.13891>.
- Anderson, M., & Leigh, A. S. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4).
- Anonymous (2018). *Initial code of conduct for data-driven health and care technology*. Department of Health & Social Care (ed) Published online 5 September 2018 ed. Her Majesty's Stationery Office.
- Antonio, V. (2018). *How AI is changing sales*. <https://hbr.org/2018/07/how-ai-is-changing-sales> Accessed 01.10.18.
- Arlitsch, K., & Newell, B. (2017). Thriving in the age of accelerations: A brief look at the societal effects of artificial intelligence and the opportunities for libraries. *Journal of Library Administration*, 57(7), 789–798.
- Autor, D. H. (2013). The 'Task Approach' to Labor Markets – An Overview. *Journal for Labour Market Research*, 46(3), 185–199.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ..., & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59.
- Bader, J., Edwards, J., Harris-Jones, C., & Hannaford, D. (1988). Practical engineering of knowledge-based systems. *Information and Software Technology*, 30(5), 266–277.
- Baek, T. H., & Morimoto, M. (2012). Stay away from me: Examining the determinants of consumer avoidance of personalized advertising. *Journal of Advertising*, 41(1), 59–76.
- Baldassarre, G., Santucci, V. G., Cartoni, E., & Caligiore, D. (2017). The architecture challenge: Future artificial-intelligence systems will require sophisticated architectures, and knowledge of the brain might guide their construction. *The Behavioral and Brain Sciences*, 40, e254.
- Baller, S., Dutta, S., & Lanvin, B. (2016). *The global information technology report 2016*. Geneva: <https://doi.org/10.1016/B978-0-12-804704-0.00010-4>.
- Barki, H., Rivard, S., & Talbot, J. (1993). A keyword classification scheme for IS research literature: An update. *MIS Quarterly*, 17(2), 209–226.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact by Solon Barocas, Andrew D. Selbst: SSRN. *California Law Review*, 104, 671–732.
- Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*, 90(10), 78–83.
- BBC (2019). *Will AI kill developing world growth*. Accessed June 2019. <https://www.bbc.co.uk/news/business-47852589>.
- Becker, G. S. (1996). *Accounting for tastes*. Cambridge: Harvard University Press.
- Bell, A. (2019). *Waiting on hold will soon become a thing of the past*. Accessed February 2019: <https://whatsnext.nuance.com/customer-experience/artificial-intelligence-bridges-gaps-between-consumer-demands-and-contact-centers/>.
- Beregi, J., Zins, M., Masson, J., Cart, P., Bartoli, J., Silberman, B., ..., & Meder, J. (2018). Radiology and artificial intelligence: An opportunity for our specialty. *Diagnostic and Interventional Imaging*, 99(11), 677–678.
- Bernard Marr. (2018). Retrieved from <https://www.forbes.com/sites/bernardmarr/2018/06/25/the-ai-skills-crisis-and-how-to-close-the-gap/#5365a24c31f3>.
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271.
- Bielenberg, A., Helm, L., Gentilucci, A., Stefanescu, D., & Zhang, H. (2012). The growth of diaspora – A decentralized online social network in the wild. *2012 proceedings IEEE INFOCOM workshops* (pp. 13–18). (March).
- Blair, W. M. (1956). *Nixon Foresees 4-Day Work Week; Says G.O.P. Policies Assure Fuller Life for Family*. *New York Times*, Sep.
- Bleicher, A. (2017). *Demystifying the black box that is AI*. Scientific American Retrieved from <https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/>.
- Bleier, A., & Eisenbeiss, M. (2015). The importance of trust for personalized online advertising. *Journal of Retailing*, 91(3), 390–409.
- Blili, S., & Raymond, L. (1993). Information technology: Threats and opportunities for small and medium-sized enterprises. *International Journal of Information Management*, 13(6), 439–448.
- Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, 46(3), 363–376.
- Bole, U., Popović, A., Žabkar, J., Papa, G., & Jaklič, J. (2015). A case analysis of embryonic data mining success. *International Journal of Information Management*, 35(2), 253–259.
- Booth, G. (2015). I, for one, welcome our new computer overlords. *IEEE Software*, 32(6), 8–10.
- Bostrom, N., & Yudkowsky, E. (2011). The ethics of Artificial Intelligence. In K. Frankish (Ed.), *Cambridge handbook of artificial intelligence*. Cambridge: Cambridge University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.
- Bradbury, D. (2018). *Does your company need an AI council?* Retrieved from <https://www.servicenow.com/workflow/corporate-ai-projects.html>.
- Breuninger, S., & Berg, M. (2001). *Great Bubbles: Reactions to the South Sea Bubble the Mississippi Scheme and the Tulip Mania Affair*.
- Brookings Institute (2019). *How artificial intelligence will affect the future of energy and climate*. Accessed 05.07.19. <https://www.brookings.edu/research/how-artificial-intelligence-will-affect-the-future-of-energy-and-climate/>.
- Brooks, R. (2002). *Robot: The future of flesh and machines*. MIT Press.
- Brooks, R. (2019). *AGI has been delayed*. Rodneybrooks.com. <https://rodneybrooks.com/agi-has-been-delayed/>.
- Brown, M. M. (2015). Revisiting the IT productivity paradox. *The American Review of Public Administration*, 45(5), 565–583.
- Brown, N., & Michael, M. (2004). Risky creatures: Institutional species boundary change in biotechnology regulation. *Health, Risk & Society*, 6(3), 207–222.
- Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *The Journal of Economic Perspectives*, 14(4), 23–48.
- Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI frontier: Modeling the global economic impact of AI*. McKinsey Global Institute 1–64 September (September). Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on fairness, accountability and transparency*, 77–91 (January).
- Burger, A. S. (2015). *Extreme working hours in western Europe and North America: A new aspect of polarization*. LEQS Paper, no. 92.
- Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Policy*, 23(1), 3–28.
- BSI. (2019) SO/IEC JTC 1/SC 42 N 222, ISO/IEC NP 23894 Information Technology – Artificial Intelligence – Risk Management <https://standardsdevelopment.bsigroup.com/projects/9018-02222> accessed on 02.06.2019.
- Calo, R. (2017). Artificial Intelligence policy: A roadmap. *SSRN Electronic Journal*, 1–28. <https://doi.org/10.2139/ssrn.3015350>.
- Capgemini Report (2018a). *Understanding digital mastery today* Capgemini Digital Transformation Institute Available from [https://www.capgemini.com/wp-content/uploads/2018/07/Digital-Mastery-DTI-report\\_20180704\\_web.pdf](https://www.capgemini.com/wp-content/uploads/2018/07/Digital-Mastery-DTI-report_20180704_web.pdf). Accessed 17.03.19.
- Capgemini Report (2018b). *TechnoVision 2018: The impact of AI* Available at <https://www.capgemini.com/technovision-2018-the-impact-of-ai/>. Accessed 17.03.19.
- Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., & Zdeborová, L. (2019). *Machine learning and the physical sciences*. arXiv:1903.10563 [physics.comp-ph]. <https://inspirehep.net/search?p=find+eprint+1903.10563>.
- Carrasquilla, J., & Melko, R. (2017). Machine learning phases of matter. *Nature Physics*, 13, 431–434.

- Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). *Exploring neural networks with activation atlases*. <https://doi.org/10.23915/distill.00015>.
- Chakraborty, A., & Kar, A. K. (2016). A review of bio-inspired computing methods and potential applications. *Proceedings of the international conference on signal, networks, computing, and systems*. New Delhi: Springer155–161.
- Chakraborty, A., & Kar, A. K. (2017). *Swarm intelligence: A review of algorithms. Nature-inspired computing and optimization*. Cham: Springer475–494.
- Chandler, D., Levitt, S. D., & List, J. A. (2011). Predicting and preventing shootings among at-risk youth. *The American Economic Review*, 101(3), 288–292.
- Chaudhri, V. K., Lane, H. C., Gunning, D., & Roschelle, J. (2013). Applications of artificial intelligence to contemporary and emerging educational challenges. *Artificial Intelligence Magazine, Intelligent Learning Technologies: Part*, 2(34), 4.
- Chaudhuri, A., & De, K. (2011). Fuzzy support vector machine for bankruptcy prediction. *Applied Soft Computing Journal*, 11(2), 2472–2486.
- Chen, J., & Stallaert, J. (2014). An economic analysis of online advertising using behavioral targeting. *MIS Quarterly*, 38(2), 429–449.
- Cheshire, W. P. (2017). Loophole: A limitation of medical artificial intelligence. *Ethics and Medicine*, 33(1), 7–12.
- Clarkson, J., Coleman, R., Keates, S., & Lebbon, C. (Eds.). (2003). *Inclusive design – Design for the whole population*. London: Springer-Verlag.
- Cleophas, T. J., & Cleophas, T. F. (2010). Artificial intelligence for diagnostic purposes: Principles, procedures and limitations. *Clinical Chemistry and Laboratory Medicine*, 48(2), 159–165.
- Cohen, J. L., & Kharas, H. (2018). *Using big data and artificial intelligence to accelerate global development*. Brookings Institution Accessed 04.07.19. <https://www.brookings.edu/research/using-big-data-and-artificial-intelligence-to-accelerate-global-development/>.
- Collingridge, D. (1980). *The social control of technology*. London: Frances Pinter.
- Collins, H. (2018). *Artificial intelligence: Against humanity's surrender to computers*. Medford: Polity.
- Columbus, L. (2016). *Ten ways big data is revolutionizing marketing and sales*. Accessed August 2018. Available from: <https://www.forbes.com/sites/louisacolumbus/2016/05/09/ten-ways-big-data-is-revolutionizing-marketing-and-sales/#1bfc272b21cf>.
- Combi, C. (2017). Editorial from the new editor-in-chief: Artificial intelligence in medicine and the forthcoming challenges. *Artificial Intelligence in Medicine*, 76, 37–39.
- Crevier, D. (1993). *AI: The tumultuous history of the search for artificial intelligence*. New York: Basic Books.
- Data Gov (2018). *Estimated number of enterprises in different states/UTs*. Retrieved from <https://data.gov.in/catalog/estimated-number-enterprises-different-statesuts>.
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Rethinking work in the age of AI*. Harvard Business Press.
- Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines*. New York: Harper Business.
- Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard Business Review*, 90(5), 70–76.
- Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review*, 96(1), 108–116.
- Autor, David H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3–30.
- De Stefano, V. (2018). 'Negotiating the Algorithm': Automation, artificial intelligence and labour protection.
- DeBrusk, C. (2018). *The risk of machine-learning bias (and how to prevent it)*. Available at <https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/>. Accessed 17.03.19.
- Desouza, K. C. (2018). *Delivering artificial intelligence in government: Challenges and opportunities*. Washington, DC: IBM Center for The Business of Government [http://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government\\_0.pdf](http://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government_0.pdf).
- Dey, D., & Kumar, S. (2010). Reassessing data quality for information products. *Management Science*, 56(12), 2316–2322.
- Dickson, B. (2019). *Explainable AI: Viewing the world through the eyes of neural networks*. Available at <https://bdtchats.com/2019/02/04/explainable-ai-gan-dissection-ibm-mit/> (accessed 21.03.19).
- DIN & DKE (2018). *German standardization roadmap industrie 4.0. V3*. Accessed 25.04.19. <https://www.din.de/blob/65354/57218767bd6da1927b181b9f2a0d5b39/roadmap-i4-0-e-data.pdf>.
- Dreyer, K., & Allen, B. (2018). Artificial intelligence in health care: Brave new world or golden opportunity? *Journal of the American College of Radiology*, 15(4), 655–657.
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data – Evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71.
- Dutton, T. (2018). *An overview of national AI strategies*. Retrieved from <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edf>.
- Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, 21(3), 719–734.
- Dwivedi, Y. K., Rana, N. P., Janssen, M., Lal, B., Williams, M. D., & Clement, M. (2017). An empirical validation of a unified model of electronic government adoption (UMEGA). *Government Information Quarterly*, 34(2), 211–230.
- Dwivedi, Y. K., Kapoor, K. K., & Chen, H. (2015a). Social media marketing and advertising. *The Marketing Review*, 15(3), 289–309.
- Dwivedi, Y. K., Wastell, D., Laumer, S., Henriksen, H. Z., Myers, M. D., Bunker, D., ..., & Srivastava, S. C. (2015b). Research on information systems failures and successes: Status update and future directions. *Information Systems Frontiers*, 17(1), 143–157.
- Dwivedi, Y. K., & Kuljis, J. (2008). Profile of IS research published in the European Journal of Information Systems. *European Journal of Information Systems*, 17(6), 678–693.
- Eden, A. H., Steinhart, E., Pearce, D., & Moor, J. H. (2012). *Singularity hypotheses: An overview. Singularity hypotheses*. Berlin Heidelberg: Springer1–12.
- Edwards, J. S., Duan, Y., & Robins, P. (2000). An analysis of expert systems for business decision making at different levels and in different roles. *European Journal of Information Systems*, 9(1), 36–46.
- Edwards, S. D. (2018). The HeartMath coherence model: Implications and challenges for artificial intelligence and robotics. *AI and Society*, 1–7. <https://doi.org/10.1007/s00146-018-0834-8>.
- Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017). AI augmented government: using cognitive technologies to redesign public sector work. *Deloitte Center for Government Insights*.
- Emmett, R. (2000). *Great bubbles: Reactions to the South Sea Bubble, the Mississippi scheme and the tulip mania affair*. London: Pickering & Chatto.
- Erikson, H., & Salzman-Erikson, M. (2016). Future challenges of robotics and artificial intelligence in nursing: What can we learn from monsters in popular culture? *The Permanente Journal*, 20(3).
- Eubanks, V. (2018). *Automating inequality*. St. Martin's Press.
- European Parliament (2017). *Robots and artificial intelligence: MEPs call for EU-wide liability rules [Plenary session. Press release]*. European Parliament News Retrieved from <http://www.europarl.europa.eu/news/en/news-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules>. (February).
- Fessler, L. (2017). *We tested bots like Siri and Alexa to see who would stand up to sexual harassment*. Quartz Magazine.
- Fleck, J. (1982). Development and establishment in artificial intelligence. In N. Elias, H. Martins, & R. Whitley (Eds.). *Scientific establishments and hierarchies* (pp. 169–217). Dordrecht: D. Reidel.
- Forbes (2019a). *Walmart unveils a new lab store that uses AI*. Accessed 10.06.19. <https://www.forbes.com/sites/walterloeb/2019/04/29/walmart-unveils-a-new-lab-store-for-the-future/#f06ea9f504f8>.
- Forbes (2019b). *Artificial intelligence, China and the U.S. – How the U.S. is losing the technology war*. Accessed 05.07.19. <https://www.forbes.com/sites/steveandriole/2018/11/09/artificial-intelligence-china-and-the-us-how-the-us-is-losing-the-technology-war/#2dcacac6d195>.
- Foy, K. (2018). *Artificial intelligence system uses transparent, human-like reasoning to solve problems*. MIT News Available from <http://news.mit.edu/2018/mit-lincoln-laboratory-ai-system-solves-problems-through-human-reasoning-0911>. Accessed 17.03.19.
- Frase, P. (2016). *Four futures: Life after capitalism*. Verso books.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.
- Fry, H. (2018). *Hello world: How to be human in the age of the machine*. London, UK: Transworld Publishers.
- Special Report - AI & Robotics Accessed, 20th June 2019. <https://www.ft.com/reports/ai-robotics>.
- Future of Life Institute (2017). *Asilomar AI principles*. Future of Life Institute Retrieved from <https://futureoflife.org/ai-principles/>.
- Galanos, V., et al. (2018). Artificial intelligence does not exist: Lessons from shared cognition and the opposition to the nature/nurture divide. In Kreps (Ed.). *13th IFIP TC 9 International Conference on Human Choice and Computers, HCC13 2018*. Held at the 24th IFIP World Computer Congress, WCC 2018.
- Galanos, V. (2019). Exploring expanding expertise: Artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018. *Technology Analysis & Strategic Management*, 31(4), 421–432.
- Gaon, A., & Stedman, I. (2019). A call to action: Moving forward with the governance of artificial intelligence in Canada. *Alberta Law Review*, 56(4).
- Garber, P. (2000). *Famous first bubbles: The fundamentals of early manias*. Cambridge: MIT Press.
- Genz, S., Lehmer, F., & Janser, M. (2018). The impact of investments in new digital technologies on wages – Worker-level evidence from Germany? *Manuscript presented at SES*.
- Gerbert, P., Reeves, M., Ransbotham, S., Kiron, D., & Spira, M. (2018). Global competition with AI in business: How China differs. *MIT Sloan Management Review* (July).
- Giannetti, C., Lucini, B., & Vadacchino, D. (2018). *Machine learning as a universal tool for quantitative investigations of phase transitions*. arXiv:1812.06726 [cond-mat.stat-mech]. <https://arxiv.org/abs/1812.06726>.
- Glauner, P., Meira, J. A., Valtchev, P., State, R., & Bettinger, F. (2017). The challenge of non-technical loss detection using artificial intelligence: A survey. *International Journal of Computational Intelligence Systems*, 10(1), 760–775.
- Goldsmith, S., & Crawford, S. (2014). *The responsive city: Engaging communities through data-smart governance*. John Wiley & Sons.
- Griffin, A. (2017). *Facebook's artificial intelligence robots shut down after they start talking to each other in their own language. The independent*. Accessed 12.05.18. Retrieved from <https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html>. (July).
- Grover, P., & Kar, A. K. (2017). Big data analytics: A review on theoretical contributions and tools used in literature. *Global Journal of Flexible Systems Management*, 18(3), 203–229.
- Guest, D., Cranmer, K., & Whiteson, D. (2018). Deep learning and its application to LHC physics. *Annual Review of Nuclear and Particle Science*, 68, 161–181.
- Gupta, R. K., & Kumari, R. (2017). Artificial intelligence in public health: Opportunities and challenges. *JK Science*, 19(4), 191–192.
- Gutierrez, A., O'Leary, S., Rana, N. P., Dwivedi, Y. K., & Calle, T. (2019). Using privacy calculus theory to explore entrepreneurial directions in mobile location-based advertising: Identifying intrusiveness as the critical risk factor. *Computers in Human Behavior*, 95, 295–306.
- Haddow, G., Bruce, A., Calvert, J., Harmon, S. H., & Marsden, W. (2010). Not “human” enough to be human but not “animal” enough to be animal – The case of the HFEA, cybrids and xenotransplantation in the UK. *New Genetics Society*, 29(1), 3–17.
- Haefner, M., & Panuwatwanich, K. (2017). *Perceived Impacts of Industry 4.0 on Manufacturing Industry and Its Workforce: Case of Germany. International conference on engineering, project, and product management*. Cham: Springer199–208 (September).
- Hamaguchi, N., & Kondo, K. (2018). *Regional employment and artificial intelligence in*



- Japan. Research Institute of Economy, Trade and Industry (RIETI).
- Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction research: Past, present, and future. *Ergonomics in Design*, 21(2), 9–14.
- Harhoff, Dietmar, Heumann, Stefan, Jentszsch, Nicola, Lorenz, Philippe Outline for a German Strategy for Artificial Intelligence (July 25, 2018). Available at SSRN: <https://ssrn.com>.
- Hays, J., & Efros, A. A. (2007). Scene completion using millions of photographs. *ACM Transactions on Graphics (TOG)*, 26(3), 4.
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72–80.
- Hilovská, K., & Koncz, P. (2012). Application of artificial intelligence and data mining techniques to financial markets. *ACTA VSFS*, 6, 62–77.
- HLSCAI (House of Lords. Select Committee on Artificial Intelligence). (2018). AI in the UK: Ready, willing, and able? Report of session 2017–19. April 16. The Authority of the House of Lords. Accessed 16.04.18. <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.
- Holak, B. (2018). *Forrester 5 AI predictions for 2019: Pragmatic AI takes hold*. Available at [https://searchcio.techtarget.com/news/252453560/5-AI-predictions-for-2019-Pragmatic-AI-takes-hold?src=5828756&asrc=EM\\_ERU\\_104394864&utm\\_content=eru-rd2-rcpF&utm\\_medium=EM&utm\\_source=ERU&utm\\_campaign=20181203\\_ERU%20Transmission%20for%2012/03/2018%20\(UserUniverse:%20466834\) \(accessed 22.03.19\)](https://searchcio.techtarget.com/news/252453560/5-AI-predictions-for-2019-Pragmatic-AI-takes-hold?src=5828756&asrc=EM_ERU_104394864&utm_content=eru-rd2-rcpF&utm_medium=EM&utm_source=ERU&utm_campaign=20181203_ERU%20Transmission%20for%2012/03/2018%20(UserUniverse:%20466834) (accessed 22.03.19)).
- Houssami, N., Lee, C. I., Buist, D. S. M., & Tao, D. (2017). Artificial intelligence for breast cancer screening: Opportunity or hype? *Breast*, 36, 31–33.
- Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172.
- Hughes, A., & Morton, M. S. S. (2006). The Transforming power of complementary assets. *MIT Sloan Management Review*, 47(4), 50–58.
- Hughes, L., Dwivedi, Y. K., Misra, S. K., Rana, N. P., Raghavan, V., & Akella, V. (2019). Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda. *International Journal of Information Management*, 49, 114–129.
- Hughes, D. L., Dwivedi, Y. K., & Rana, N. P. (2017). Mapping IS failure factors on PRINCE2® stages: An application of interpretive ranking process (IRP). *Production Planning & Control*, 28(9), 776–790.
- Hughes, D. L., Dwivedi, Y. K., Rana, N. P., & Simintiras, A. C. (2016). Information systems project failure – Analysis of causal links using interpretive structural modelling. *Production Planning & Control*, 27(16), 1313–1333.
- Hughes, L., Wang, X., & Chen, T. (2012). A review of protocol implementations and energy efficient cross-layer design for wireless body area networks. *Sensors*, 12(11), 14730–14773.
- Hurley, K. (2017). *11 artificial intelligence tools, transforming the B2B sales world*. <https://nudge.ai/artificial-intelligence-sales-tools/>. Accessed 01.10.18.
- IDC (2019). *Automation and Customer Experience Needs Will Drive AI Investment to \$5 Billion by 2019 Across European Industries*. Accessed 01.08.19. <https://www.idc.com/getdoc.jsp?containerId=prEMEA44978619>.
- Ilavarasan, P. V. (2018). Automation and workforce in India: Terrible consequences or impossible? In H. Galperin, & A. Alarcon (Eds.). *The future of work in the global south* (pp. 16–21). International Development Research Centre.
- Artificial Intelligence: The Public Policy Opportunity. Accessed on July 25th 2019. <https://www.intel.ai/solutions/artificial-intelligence-the-public-policy-opportunity/#gs.wu66jy>.
- Ismagilova, E., Hughes, L., Dwivedi, Y. K., & Raman, K. R. (2019). Smart cities: Advances in research—An information systems perspective. *International Journal of Information Management*, 47, 88–100.
- ITU (2017). *Measuring the information society report 2017: Volume 1*. Geneva.
- Jai, T. M., Burns, L. D., & King, N. J. (2013). The effect of behavioral tracking practices on consumers' shopping evaluations and repurchase intention toward trusted online retailers. *Computers in Human Behavior*, 29(3), 901–909.
- Jain, P. K., & Mosier, C. T. (1992). Artificial intelligence in flexible manufacturing systems. *International Journal of Computer Integrated Manufacturing*, 5(6), 378–384.
- Janssen, M., & Kuk, G. (2016a). Big and Open Linked Data (BOLD) in research, policy and practice. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), 3–13.
- Janssen, M., & Kuk, G. (2016b). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371–377.
- Janssen, M., Matheus, R., & Zuiderwijk, A. (2015). Big and Open Linked Data (BOLD) to create smart cities and citizens: Insights from smart energy and mobility cases. *Paper presented at the Proceedings of the 14th IFIP Electronic Government (EGOV) conferences*.
- Janssen, M., Van de Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70(1), 338–345.
- Jarrah, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586.
- Jonsson, A., & Svensson, V. (2016). *Systematic lead time analysis*. Chalmers University of Technology Accessed April 2019. <http://www.publications.lib.chalmers.se/records/fulltext/238746/238746.pdf>.
- Joy, B. (2000). Why the future doesn't need us. *Wired Magazine*, 8(4), 238–262.
- Juniper Research (2018). *AI in retail. segment analysis, vendor positioning & market forecasts 2019–2023*. Accessed June 2019. <https://www.juniperresearch.com/researchstore/fintech-payments/ai-in-retail>.
- Kahn, C. E. (2017). From images to actions: Opportunities for artificial intelligence in radiology. *Radiology*, 285(3), 719–720.
- Kahneman, D. (2011). *Thinking, fast and slow*. London, UK: Macmillan.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292.
- Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? Improving public policy by predicting hygiene inspections using online reviews. *Proceedings of the 2013 conference on empirical methods in natural language processing*, 1443–1448.
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers*, 20(3), 531–558.
- Kar, A. K. (2016). Bio inspired computing – A review of algorithms and scope of applications. *Expert Systems with Applications*, 59, 20–32.
- Katz, L. F., & Margo, R. A. (2013). *Technical change and the relative demand for skilled labor: The United States in historical perspective*. Tech. Rep., NBER Working Paper No. 18752, National Bureau of Economic Research.
- Katz, Y. (2017). *Manufacturing an artificial intelligence revolution*. Available at SSRN 3078224.
- Keynes, J. M. (2010). *Economic possibilities for our grandchildren. Essays in persuasion*. London: Palgrave Macmillan 321–332.
- Khanna, S., Sattar, A., & Hansen, D. (2013). Artificial intelligence in health – The three big challenges. *Australasian Medical Journal*, 6(5), 315–317.
- Klaus Schwab (2016). *The fourth industrial revolution*. World Economic Forum 2016.
- Kling, R. (1991). Computerization and social transformations. *Science, Technology, & Human Values*, 16(3), 342–367.
- Knight, W. (2017). *"The dark secret at the heart of AI." Intelligent Machines*. <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>. Accessed 05.04.19.
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210.
- Krishen, A. S., Raschke, R. L., Close, A. G., & Kachroo, P. (2017). A power-responsibility equilibrium framework for fairness: Understanding consumers' implicit privacy concerns for location-based services. *Journal of Business Research*, 73, 20–29.
- Kumar, S. L. (2017). State of the art-intense review on artificial intelligence systems application in process planning and manufacturing. *Engineering Applications of Artificial Intelligence*, 65, 294–329.
- Kumar, V., & Shaphali, G. (2016). Conceptualizing the evolution and future of advertising. *Journal of Advertising*, 45(3), 302–317.
- Kusiak, A. (1987). Artificial intelligence and operations research in flexible manufacturing systems. *INFOR: Information Systems and Operational Research*, 25(1), 2–12.
- Kwong, C. K., Jiang, H., & Luo, X. (2016). AI-based methodology of integrating additive design, engineering, and marketing for defining design specifications of new products. *Engineering Applications of Artificial Intelligence*, 47, 49–60.
- Langton, J. (1984). The industrial revolution and the regional geography of England. *Transactions of the Institute of British Geographers*, 9(2), 145–167.
- Lassau, N., Estienne, T., de Vomecourt, P., Azoulay, M., Cagnol, J., Garcia, G., ..., & Cotten, A. (2019). Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. *Diagnostic and Interventional Imaging*, 100(4), 199–209.
- Laudon, C. K., & Laudon, J. P. (2017). *Management information systems: Managing the digital firm* (15th ed.). New Delhi: Pearson Education India.
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293–303.
- Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4), 431–440.
- Lee, J. H. (2002). Artificial intelligence-based sampling planning system for dynamic manufacturing process. *Expert Systems with Applications*, 22(2), 117–133.
- Li, B. H., Hou, B. C., Yu, W. T., Lu, X. B., & Yang, C. W. (2017). Applications of artificial intelligence in intelligent manufacturing: A review. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 86–96.
- Linden, A., & Fenn, J. (2003). *Understanding Gartner's hype cycles. Strategic Analysis Report N° R-20-1971*. Gartner, Inc.
- Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013). The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility. *Decision Support Systems*, 54(3), 1452–1462.
- Liu, J., Qi, Y., Yang Meng, Z., & Fu, L. (2017). Self-learning Monte Carlo method. *Physical Review B*, 95. <https://doi.org/10.1103/PhysRevB.95.041101> 041101(R).
- Liu, Y., Chan, C., Zhao, C., & Liu, C. (published online 2018). Unpacking knowledge management practices in China: Do institution, national and organizational culture matter? *Journal of Knowledge Management*. <https://doi.org/10.1108/JKM-07-2017-0260>.
- Löffler, M., & Tschiesner, A. (2013). *The Internet of things and the future of manufacturing*. McKinsey & Company Accessed in April 2019. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-internet-of-things-and-the-future-of-manufacturing>.
- London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Center Report*, 49, 15–21.
- Loring, Evan (2018). *How AI will help sales representatives*. <https://readwrite.com/2018/09/27/how-ai-will-help-sales-representatives/>. Accessed 01.10.18.
- Loten, Angus (2017). *AI to drive job growth by 2020: Gartner*. <https://blogs.wsj.com/cio/2017/12/15/ai-to-drive-job-growth-by-2020-gartner/>. Accessed 10.01.18.
- Ma, A. (2018). A psychological ailment called 'hikikomori' is imprisoning 500,000 Japanese people in their homes — and it's more of a threat than ever. *Business Insider*.
- MacKenzie, D. (1999). The certainty trough. In R. Williams, W. Faulkner, & J. Fleck (Eds.). *Exploring expertise: Issues and perspectives*. London: MacMillan Press.
- Macknik, S. L., & Martinez-Conde, S. (2011). *Sleights of mind*. Surrey, UK: Picador.
- Mauro, K. (2019). A human-centric perspective exploring the readiness towards smart warehousing: The case of a large retail distribution warehouse. *International Journal of Information Management*, 45, 176–190.
- Makridakis, S. (2018). Forecasting the impact of artificial intelligence, Part 3 of 4: The potential effects of AI on businesses, manufacturing, and commerce. *Foresight: The International Journal of Applied Forecasting*, 49, 18–27.
- Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., ..., & Sanghvi, S. (2017). *Jobs lost, jobs gained: Workforce transitions in a time of automation*. McKinsey Global Institute.

- Markus, M. L. (2017). Datification, organizational strategy, and IS research: What's the score? *Journal of Strategic Information Systems*, 26, 233–241.
- McBride, S. E., Rogers, W. A., & Fisk, A. D. (2014). Understanding human management of automation errors. *Theoretical Issues in Ergonomics Science*, 15(6), 545–577.
- McKinsey Global Institute (2017). *Future that works: Automation, employment, and productivity*. New York: McKinsey Global Institute.
- McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Medaglia, R., & Zhu, D. (2017). Public deliberation on government-managed social media: A study on Weibo users in China. *Government Information Quarterly*, 34(3), 533–544.
- Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Cambridge, MA, USA: Harvard University Press.
- METI – Ministry of Economy, T. and I (2016). *Vision of new industrial structure – Japan's strategies for taking the lead in the fourth industrial revolution – Interim Report by New Industrial Structure Committee*. Japan: METI Retrieved from [http://www.meti.go.jp/english/policy/economy/industrial\\_council/pdf/innovation160427a.pdf](http://www.meti.go.jp/english/policy/economy/industrial_council/pdf/innovation160427a.pdf).
- Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), <https://doi.org/10.1098/rsta.2017.0357>.
- Milano, M., O'Sullivan, B., & Gavanelli, M. (2014). Sustainable policy making: A strategic challenge for artificial intelligence. *AI Magazine*, 35(3), 22–35.
- Miller, S. (2018). AI: Augmentation, more so than automation. *Asian Management Insights*, 5(1), 1–20.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38.
- Mitchell, M. (2019). Artificial intelligence hits the barrier of meaning. *Information (Switzerland)*, 10(2), <https://doi.org/10.3390/info10020051>.
- Mogaji, E., Olaleye, S., & Ukpabi, D. (2018). *Using AI to personalise emotionally appealing advertisement*. Swansea, Wales.
- Moglen, E. (2013). Privacy and security the tangled web we have woven. *Communications of the ACM*, 56(2), 20–22.
- Moniz, A. B. (2013). *Robots and humans as co-workers?* IET Working Papers Series, WPS03. Retrieved from <http://arxiv.org/abs/1507.06136>.
- Morabito, R., Cozzolino, V., Ding, A. Y., Beijar, N., & Ott, J. (2018). Consolidate IoT edge computing with lightweight virtualization. *IEEE Network*, 32(1), 102–111.
- Morikawa, M. (2017). *Who are afraid of losing their jobs to artificial intelligence and robots? Evidence from a survey*, GLO Discussion Paper, No. 71. Maastricht: Global Labor Organization (GLO).
- Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman, & M. Mouloua (Eds.). *Automation and human performance: Theory and application* (pp. 201–220). Mahwah, NJ: Erlbaum.
- Muhuri, P. K., Shukla, A. K., & Abraham, A. (2019). Industry 4.0: A bibliometric analysis and detailed overview. *Engineering Applications of Artificial Intelligence*, 78, 218–235.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Müller, V. C., & Bostrom, N. (2016). *Future progress in artificial intelligence: A survey of expert opinion*. *Fundamental issues of artificial intelligence*. Cham: Springer 555–572.
- Nambu, T. (2016). Legal regulations and public policies for next-generation robots in Japan. *AI and Society*, 31(4), 483–500. <https://doi.org/10.1007/s00146-015-0628-1>.
- Narayanan, A. (2013). Society under threat .. but not from AI. *AI and Society*, 28(1), 87–94.
- Negnevitsky, M. (2011). *Artificial intelligence – A guide to intelligent systems*. Book. <https://doi.org/10.1145/2063176.2063177>.
- Nguyen, G. K., & Shetty, A. S. (2018). Artificial intelligence and machine learning: Opportunities for radiologists in training. *Journal of the American College of Radiology*, 15(9), 1320–1321.
- Nikolic, B., Ignjatovic, J., Suzic, N., Stevanov, B., & Rikalovic, A. (2017). Predictive manufacturing systems in industry 4.0: Trends, benefits and challenges. *Annals of DAAAM & Proceedings*, 28.
- Niti Aayog, I. (2018). *National strategy for artificial intelligence*. Discussion Paper. (September).
- O'Neil, C. (2016). *Weapons of math destructions*. New York: Crown.
- Olanrewaju, A. S. T., Hossain, M. A., Whiteside, N., & Mercieca, P. (2020). Social media and entrepreneurship research: A literature review. *International Journal of Information Management*, 50, 90–110.
- Olshannikova, E., Ometov, A., Koucheryav, Y., & Olsson, T. (2015). Visualizing big data with augmented and virtual reality: Challenges and research agenda. *Journal of Big Data*, 2(1), <https://doi.org/10.1186/s40537-015-0031-2>.
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation. *Human Factors*, 56(3), 476–488.
- Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: Paving the way towards digital transformation and sustainable societies. *Information Systems and e-Business Management*, 16(3), 479–491.
- Parasuraman, R. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics*, 43(7), 931–951.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410.
- Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, 50(3), 511–520.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3), 286–297.
- Parveen, R. (2018). Artificial intelligence in construction industry: Legal issues and regulatory challenges. *International Journal of Civil Engineering and Technology*, 9(13), 957–962.
- Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
- Patsiotis, A., Hughes, T., & Webber, D. (2012). Adopters and non-adopters of Internet banking: A segmentation study. *International Journal of Bank Marketing*, 30(1), 20–42.
- Pearl, J., & MacKenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, USA: Basic Books.
- Pickering, A. (2009). Cybernetics as nomad science. *Deleuzian intersections in science. Technology and Anthropology*, 155–162.
- Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*, 6(1).
- Pratap, K. V. (2015). *Financing of smart cities*. Retrieved from <http://smartcities.gov.in/upload/uploadfiles/files/Financing-of-Smart-Cities.pdf>.
- Preece, A. (2018). Asking 'why' in AI: Explainability of intelligent systems – Perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, 25, 63–72.
- Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT Sloan Management Review*, 59(1).
- Raval (2018). AI, IoT, blockchain enhance efficiency of Akshaya Patra's Mid-Day Meal Program. *Express Computer* July 25, 2018. <https://www.expresscomputer.in/news/ai-iot-blockchain-enhance-efficiency-of-akshaya-patras-mid-day-meal-program/21146/>.
- Reza Tizhoosh, H., & Pantanowitz, L. (2018). Artificial intelligence and digital pathology: Challenges and opportunities. *Journal of Pathology Informatics*, 9(1).
- Risse, M. (2019). Human rights and artificial intelligence: An urgently needed agenda. *Human Rights Quarterly*, 41(1), 1–16.
- Roberts, N., Galluch, P. S., Dinger, M., & Grover, V. (2012). Absorptive capacity and information systems research: Review, synthesis, and directions for future research. *MIS Quarterly*, 36(2), 625–648.
- Robitzky, D. (2018). *Welsh police used face scanning software that incorrectly flagged thousands*. Futurism.com. Retrieved from <https://futurism.com/welsh-police-face-scanning-software-incorrect/>. (May).
- Rubik, B., & Jabs, H. (2018). Artificial intelligence and the human biofield: New opportunities and challenges. *Cosmos and History*, 14(1), 153–162.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Malaysia: Pearson Education Limited.
- Sawyer, S. (2008). Data wealth, data poverty, science and cyberinfrastructure. *Prometheus*, 26(4), 355–371.
- Scherer, M. U. (2016). Regulating artificial intelligence systems. *Harvard Journal of Law and Technology*, 29(2), 353–400.
- Schulz, P. J., & Nakamoto, K. (2013). Patient behavior and the benefits of artificial intelligence: The perils of “dangerous” literacy and illusory patient empowerment. *Patient Education and Counseling*, 92(2), 223–228.
- Scitovsky, T. (1976). *The joyless economy: An inquiry into human satisfaction and consumer dissatisfaction* Oxford. England: Oxford U Press.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
- Sebok, A., & Wickens, C. D. (2017). Implementing lumberjacks and black swans into model-based tools to support human – Automation interaction. *Human Factors*, 59(2), 189–203.
- Senyo, P. K., Liu, K., & Effah, J. (2019). Digital business ecosystem: Literature review and a framework for future research. *International Journal of Information Management*, 47, 52–64.
- Shackle, G. (1949 [1952]). *Expectation in economics*. Cambridge: Cambridge University Press.
- Shanahan, P. E., Trewartha, D., & Detmold, W. (2018). Machine learning action parameters in lattice quantum chromodynamics. *Physical Review D*, 97. <https://doi.org/10.1103/PhysRevD.97.094506>.
- Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics*, 1(1), 89–129.
- Shneiderman, B. (2016). Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, 113(48), 13538–13540. <https://doi.org/10.1073/pnas.1618211113>.
- Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., & Bauer, Z. (2018). *The AI index 2018 annual report*. AI Index Steering Committee, human-centered AI initiative Stanford University Available from [http://cdn.aiindex.org/2018/AI%20Index\\_202018](http://cdn.aiindex.org/2018/AI%20Index_202018).
- Shukla, N., Tiwari, M. K., & Beydoun, G. (2018). Next generation smart manufacturing and service systems using big data analytics. *Computers & Industrial Engineering*, 128, 905–910.
- Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2), 47–53.
- Smit, E. G., Van Noort, G., & Voorveld, H. A. (2014). Understanding online behavioural advertising: User knowledge, privacy concerns, and online coping behaviour in Europe. *Computers in Human Behavior*, 32, 15–22.
- Smith, A. (1759 [1976]). The theory of moral sentiments. In D.D. Raphael, & A. L. Mactie (Eds.), *Liberty classics*. Indianapolis: Liberty Press.
- Snow, J. (2018). We're in a diversity crisis": Cofounder of Black in AI on what's poisoning algorithms in our lives. *MIT Technology Review* (February).
- Spanaki, K., Gürgüç, Z., Adams, R., & Mulligan, C. (2018). Data supply chain (DSC): Research synthesis and future directions. *International Journal of Production Research*, 56(13), 4447–4466.
- Spencer, J., Poggi, J., & Gheerawo, R. (2018). Designing out stereotypes in artificial intelligence: Involving users in the personality design of a digital assistant. *Proceedings of the 4th EAI international conference on smart objects and technologies for social good*, 130–135.
- Stalin, J. (1952). *Economic problems of socialism in the USSR*. Moscow 36.
- Stead, W. W. (2018). Clinical implications and challenges of artificial intelligence and deep learning. *JAMA – Journal of the American Medical Association*, 320(11), 1107–1108.
- Steinmueller, W. E. (2001). ICTs and the possibilities for leapfrogging by developing



- countries. *International Labour Review*, 140(2), 193–210.
- Stone, P. (2016). *Artificial intelligence and life in 2030*. Stanford University Report, 52.
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368–383.
- Sunstein, C. R. (2017). *Republic: Divided democracy in the age of social media*. Princeton, NJ: Princeton University Press.
- Szollosy, M. (2016). Freud, Frankenstein and our fear of robots: Projection in our cultural perception of technology. *AI & Society*, 32, 433–439.
- Tadesse, B., & White, R. (2010). Cultural distance as a determinant of bilateral trade flows: Do immigrants counter the effect of cultural differences? *Applied Economics Letters*, 17(2), 147–152.
- Tamilmani, K., Rana, N. P., Prakasam, N., & Dwivedi, Y. K. (2019). The battle of brain vs. heart: A literature review and meta-analysis of “hedonic motivation” use in UTAUT2. *International Journal of Information Management*, 46, 222–235.
- Tangermann, V. (2018). *Criminals are now using swarms of small drones to befuddle law enforcement*. Retrieved from <https://futurism.com>. Website: <https://futurism.com/criminals-swarms-drones-law-enforcement/>.
- Tarafdar, M., Beath, C. M., & Ross, J. W. (2017). Enterprise cognitive computing applications – Opportunities and challenges. *IT Professional*, 19(4), 21–27.
- The Economist Intelligence Unit (2018). *Intelligent economies: AI's transformation of industries and society*. Economist.
- The Economist Intelligence Unit (2016). *Advanced science and the future of government: Robots and artificial intelligence, genomic medicine, biometrics*. Economist.
- Thesmar, D., Sraer, D., Pinheiro, L., Dadson, N., Veliche, R., & Greenberg, P. (2019). Combining the power of artificial intelligence with the richness of healthcare claims data: Opportunities and challenges. *Pharmacoeconomics*. <https://doi.org/10.1007/s40273-019-00777-6>.
- Thrall, J. H., Li, X., Li, Q., Cruz, C., Do, S., Dreyer, K., & Brink, J. (2018). Artificial intelligence and machine learning in radiology: Opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology*, 15(3), 504–508.
- Torre, A. (2008). On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies*, 42(6), 869–889.
- Tubadj, A., & Nijkamp, P. (2016). Six degrees of cultural diversity and R&D output efficiency: Cultural percolation of new ideas – An illustrative analysis of Europe. *Letters in Spatial and Resource Sciences*, 9(3), 247–264.
- Tubadj, A., & Nijkamp, P. (2015). Cultural Gravity effects among migrants: A comparative analysis of the EU15. *Economic Geography*, 91(3), 344–380.
- Tubadj, A., Angelis, V., & Nijkamp, P. (2016). Cultural hysteresis, entrepreneurship and economic crisis: An analysis of buffers to unemployment after economic shocks. *Cambridge journal of regions. Economy and Society*, 9(1), 103–136.
- Tubadj, A., Webber, D., & Denny, T. (2019). *Cultural relativity and AI adoption: Robo-advisory in retail banking*. Manuscript.
- Turkle, S. (1981). Computers as rohrschach: Subjectivity and social responsibility. In B. Sundin (Ed.), *Is the computer a tool?* (pp. 81–99). Stockholm: Almqvist and Wiksell.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- UKRI (2018). [epsrc.ukri.org/funding/calls/aicdts2018full/](https://www.ukri.org/funding/calls/aicdts2018full/).
- UKRI. (2019). [www.ukri.org/news/200m-to-create-a-new-generation-of-artificial-intelligence-leaders/](https://www.ukri.org/news/200m-to-create-a-new-generation-of-artificial-intelligence-leaders/).
- United Nations (2018). *E-government survey 2018*. New York: Department of Economic and Social Affairs.
- Vagia, M., Transeth, A. A., & Fjerdings, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, 53, 190–202.
- Van Doorn, J., & Hoekstra, J. C. (2013). Customization of online advertising: The role of intrusiveness. *Marketing Letters*, 24(4), 339–351.
- Van Est, R., Gerritsen, J., & Kool, L. (2017). *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*. – Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE) The Hague: Rathenau Instituut.
- Van Lente, H. (2012). Navigating foresight in a sea of expectations: Lessons from the sociology of expectations. *Technology Analysis & Strategic Management*, 24(8), 769–782.
- Varga-Szemes, A., Jacobs, B. E., & Schoepf, U. J. (2018). The power and limitations of machine learning and artificial intelligence in cardiac CT. *Journal of Cardiovascular Computed Tomography*, 12(3), 202–203.
- Vesset, D., Olofson, C. W., Nadkorni, A., Zaidi, A., McDonough, B., Schubmehl, D., et al. (2015). *Futurescape: Worldwide big data and analytics 2016 predictions*. IDC Retrieved from <https://www.idc.com/research/viewtoc.jsp?containerId=259835>.
- Vincent, J. (2016). *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*. The Verge Available from <https://www.theverge.com/2016/3/24/11297050/twitter-taught-microsofts-ai-chatbot-to-be-a-racist-asshole-in-less-than-a-day>.
- tay-microsoft-chatbot-racist.
- Vincent, J. (2019). *The state of AI in 2019*. Available from: <https://www.theverge.com/2019/1/28/18197520/ai-artificial-intelligence-machine-learning-computational-science> [Accessed 03.03.19].
- von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4(4), 404–409.
- Walsh, T. (2018). Expert and non-expert opinion about technological unemployment. *International Journal of Automation and Computing*, 15(5), 637–642.
- Walton, P. (2014). A model for information. *Information*, 5(3), 479–507.
- Walton, P. (2015a). Measures of information. *Information*, 6(1), 23–48.
- Walton, P. (2015b). Digital information and value. *Information*, 6(4), 733–749.
- Walton, P. (2017). Information and Inference. *Information*, 8(2), 61.
- Walton, P. (2018a). Artificial intelligence and the limitations of information. *Information (Switzerland)*, 9(12), <https://doi.org/10.3390/info9120332>.
- Walton, P. (2018b). Artificial Intelligence and the Limitations of Information. *Information*, 9(12), 332.
- Wang, L. (2016). Discovering phase transitions with unsupervised learning. *Physical Review B*, 94, 195105. <https://doi.org/10.1103/PhysRevB.94.195105>.
- Wang, L., & Wang, X. V. (2016). *Outlook of cloud, CPS and IoT in manufacturing. Cloud-based cyber-physical systems in manufacturing*. Cham: Springer 377–398.
- Wang, L., Törngren, M., & Onori, M. (2015a). Current status and advancement of cyber-physical systems in manufacturing. *Journal of Manufacturing Systems*, 37, 517–527.
- Wang, X., Li, X., & Leung, V. C. M. (2015b). Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges. *IEEE Access*, 3, 1379–1391.
- Weber, F., & Schütte, R. (2019). Domain-oriented analysis of the impact of machine learning—The case of retailing. *Big Data Cognition Computation*, 3(1), 1–14.
- Westerman, G., Bonnet, D., & McAfee, A. (2014). *Leading digital: Turning technology into business transformation*. Cambridge, MA, USA: Harvard Business Review Press.
- Wihlborg, E., Larsson, H., & Hedström, K. (2016). “The computer says no!” – A case study on automated decision-making in public authorities. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 2903–2912. <https://doi.org/10.1109/HICSS.2016.364>.
- Wilson, J., & Daugherty, P. R. (2018). Collaborative intelligence humans and AI are joining forces. *Harvard Business Review*, 96(4), 115–123.
- Wirth, N. (2018). Hello marketing, what can artificial intelligence help you with? *International Journal of Market Research*, 60(5), 435–438.
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public Sector—Applications and challenges. *International Journal of Public Administration*, 42(7), 596–615.
- Xu, J., Yang, P., Xue, S., Sharma, B., Sanchez-Martin, M., Wang, F., ..., & Parikh, B. (2019). Translating cancer genomics into precision medicine with artificial intelligence: Applications, challenges and future perspectives. *Human Genetics*. <https://doi.org/10.1007/s00439-019-01970-5>.
- Yang, J., Chen, Y., Huang, W., & Li, Y. (2017). Survey on artificial intelligence for additive manufacturing. *2017 23rd International Conference on Automation and Computing (ICAC)* (pp. 1–6). (September).
- Yim, D. S., Cho, H. H., Song, C. U., Lee, J., Lee, S., & Park, S. (2016). The strategy of technology park for the development of IT industry in Pakistan. *2016 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 176–199). <https://doi.org/10.1109/PICMET.2016.7806568>.
- Yoon, M., & Baek, J. (2016). Paideia education for learners' competencies in the age of artificial intelligence – The google DeepMind challenge match. *International Journal of Multimedia and Ubiquitous Engineering*, 11(11), 309–318.
- Zandi, D., Reis, A., Vayena, E., & Goodman, K. (2019). New ethical challenges of digital technologies, machine learning and artificial intelligence in public health: A call for papers. *Bulletin of the World Health Organization*, 97(1), 2.
- Zatarain, J. M. N. (2017). The role of automated technology in the creation of copyright works: The challenges of artificial intelligence. *International Review of Law, Computers and Technology*, 31(1), 91–104.
- Zheng, Y., Wu, W., Chen, Y., Qu, H., & Ni, L. M. (2016). Visual analytics in urban computing: An overview. *IEEE Transactions on Big Data*, 2(3), 276–296. <https://doi.org/10.1109/TBDDATA.2016.2586447>.
- Zhong, R. Y., Xu, C., Chen, C., & Huang, G. Q. (2017b). Big data analytics for physical internet-based intelligent manufacturing shop floors. *International Journal of Production Research*, 55(9), 2610–2621.
- Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017a). Intelligent manufacturing in the context of industry 4.0: A review. *Engineering*, 3(5), 616–630.
- Zlotowski, J., Yogeewaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100, 48–54.