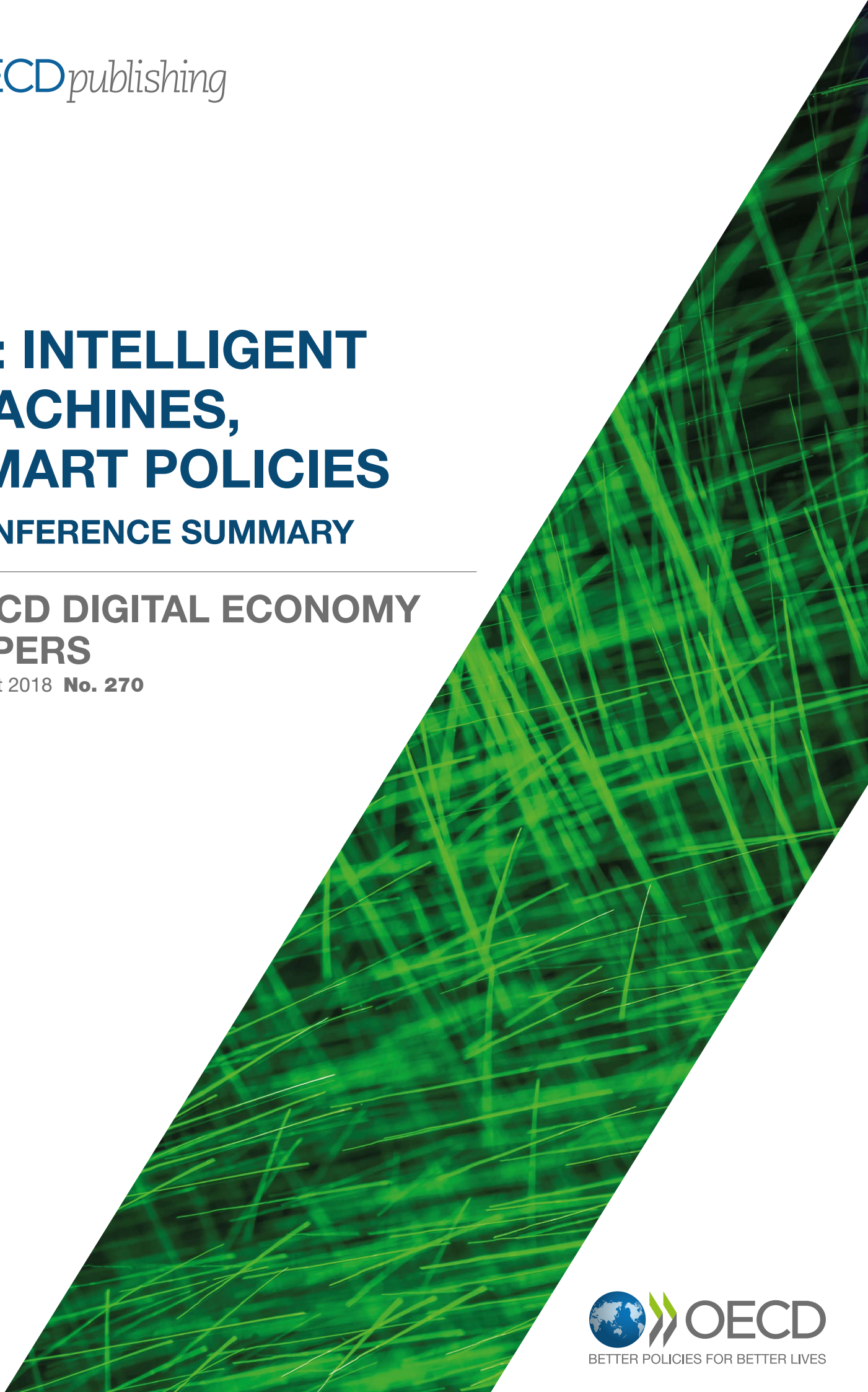
OECD publishing

AI: INTELLIGENT MACHINES, SMART POLICIES

CONFERENCE SUMMARY

OECD DIGITAL ECONOMY PAPERS

August 2018 **No. 270**



OECD
BETTER POLICIES FOR BETTER LIVES

This document contains a summary of the proceedings of the conference entitled "AI: Intelligent Machines, Smart Policies", held at the OECD headquarters in Paris, France, on 26-27 October 2017. It was approved and declassified by the Committee on Digital Economy Policy on 18 May 2018 and prepared for publication by the OECD Secretariat.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CDEP(2018)8/FINAL

Foreword

This document contains a summary of the proceedings of the conference entitled "AI: Intelligent Machines, Smart Policies", held at the OECD headquarters in Paris, France, on 26-27 October 2017. It was drafted by Karine Perset and Nobuhisa Nishigata of the OECD Secretariat, with the participation of Claire Jolly, Stijn Broecke and Anne Carblanc. Conference speakers reviewed the draft and provided input and corrections.

The event organising team included, from the division on Digital Economy Policy of the OECD Directorate for Science, Technology and Innovation (STI) – in alphabetical order – Brigitte Acoca, Sarah Ferguson, Anna-Sophie Liebender, Nobuhisa Nishigata, Elettra Ronchi, Cristina Serra-Vallejo, Yuki Yokomori, with overall coordination by Karine Perset. Claire Jolly and Alistair Nolan from STI's division on Science and Technology Policy, and Stijn Broecke of the OECD Directorate for Employment, Labour and Social Affairs co-organised the event. The contributions of Shayne MacLachlan of the OECD Public Affairs and Communication Directorate and Suzanna Grant-Kejairi of the OECD's Executive Directorate are gratefully acknowledged. Anne Carblanc, Head of the OECD Digital Economy Policy Division; Andrew Wyckoff, OECD Director for Science, Technology and Innovation and Douglas Frantz, OECD Deputy Secretary-General, provided leadership and oversight.

The conference was sponsored by the Japanese Ministry of Internal Affairs and Communications (MIC). It brought together policymakers, representatives of civil society and AI experts from industry and academia to discuss opportunities and challenges and the role of policy and international co-operation. The event attracted some 50 speakers and 300 participants. Government delegates from a range of domains participated, including digital economy ministries, labour ministries, space agency representatives, research ministries, data protection authorities, and consumer protection agencies.

Interactive demonstrations were provided by Google Arts & Culture and Facebook. Google Arts & Culture demonstrated machine learning experiments on art exhibits resulting from its collaboration with over 1 200 international museums, galleries and institutions from 70 countries, available via its online application (https://artsandculture.google.com/). Facebook demonstrated the *"Facebook 360 Innovation Tour"* showcasing its latest AI and virtual reality technologies with Samsung Gear VR powered by Oculus.

More information, including presentations and pictures, is available at http://oe.cd/ai.

Table of contents

Executive Summary

This report reflects presentations and exchanges at the OECD conference "*AI: Intelligent Machines, Smart Policies*". The conference drew 50 speakers and 300 technologists, senior policy makers and representatives of civil society, labour and business to the OECD in Paris on October 26 and 27, 2017 (http://oe.cd/ai).

The discussion focused on a coordinated policy response to the transformation of society, government and industry being driven by artificial intelligence (AI) systems. The conference objectives were to encourage a vigorous exchange among policy makers, researchers, academics and the public, to inform the future work of the OECD "Going Digital" project, and to inform OECD work on policies for artificial intelligence that will help create broadly shared prosperity and unlock new opportunities for progress in critical areas such as health, education and the environment.

AI is transforming economic and social sectors deeper and faster than expected

AI is neither science fiction nor a science project. There was universal agreement that artificial intelligence already provides beneficial applications that are used every day by people worldwide. Going forward, conference participants suggested that the development and uses of AI systems should be guided by principles that will promote well-being and prosperity while protecting individual rights and democracy.

A consensus emerged that the fast-paced and far-reaching changes from AI offer dynamic opportunities for improving the economic and social sectors. AI can make business more productive, improve government efficiency and relieve workers of mundane tasks. It can also address many of our most pressing global problems, such as climate change and wider access to quality education and healthcare.

AI is moving fast, so should governments

AI policy is an urgent concern. Speakers pointed out that AI development is at a pivotal point. The rapid pace of AI research, coupled with the speed of its real-world deployment, dramatically shrinks the time frame and the distinction between fundamental research and the impact of AI on work and play. Driven by private sector research, the leap from the lab to the office or factory floor has decreased progressively over the last five years. These factors underscore the need for a robust and timely engagement between government, industry, policy and technical experts and the public.

A recurring theme was the potential for AI to increase the efficiency and effectiveness of entire sectors, including the delivery of public services. Applied wisely, AI can improve well-being of people in areas like education, public safety, health and work-life balance. Governments need to plan for the investing in and developing of AI for its many benefits.

New policies are necessary to adapt to AI in the workplace

Another key focus was the role of AI in defining the future of work. AI is taking over some tasks long performed by humans. These changes will create new opportunities in the workplace and work-life balance. But they will also disrupt the livelihoods of millions of people. Deep learning, supported by reinforcement learning, has recently led to impressive advances in AI capacity. In areas such as image recognition, AI performance now exceeds human

performance: between 2010 and 2017, image recognition accuracy reportedly grew from less than 70% to 96%; human performance accuracy remained at 95%.

There was uncertainty about the speed and scale of the transition, but a consensus that governments should adapt existing policies and develop new strategies that prepare citizens, educators and businesses for the jobs of the future and minimise the negative impacts. An emphasis was placed on education and training for workers coming into the labour force and retraining for those displaced by AI.

In addition to the impact of automation and robotics on the workforce, many participants said governments and business should cooperate to create policies to ensure that AI does not widen the economic divides between people, companies of different sizes, countries and continents. Policies and programmes were outlined to minimise widespread economic disruption and provide broader access to AI technologies, notably to help small- and medium-sized enterprises (SMEs) navigate the AI transition. Targeted investments in R&D, access to data and upskilling were among the tools highlighted for providing broad access to AI for companies and sectors in danger of being left behind.

Similarly, participants expressed concern about the concentration of technology and financial resources in the hands of a few companies and nations. The notion of "winner-takes-most" markets was flagged as a key issue. Labour representatives worried about the impact on worker autonomy of cost-saving mechanisms and rising productivity in dominant AI companies. Large companies explained how they are trying to address these issues. For example, Google distributes its machine-learning technology and some training data sets freely.

Related concerns were voiced about the shortage of skilled workers and the difficulty of smaller firms and the public sector in competing with dominant firms for scarce talent. Without skilled workers, SMEs and public agencies risk falling behind. Several European participants said sharing the benefits of AI in Europe is restricted by the flow of researchers and engineers to other regions and the migration from the public sector to the private sector. Others said the brightest AI computer scientists are often snapped up by revenue-driven organisations with lucrative salaries rather than tackling important social challenges through research and public-sector work.

Benefiting from AI requires enhanced access to data

The dependence of AI on vast quantities of data creates complex legal, cultural and technical issues surrounding data protection, data regulation and the data economy. One of the strongest themes that emerged from the conference was the need for enhanced access to data to leverage AI broadly, through open data policies, data interoperability and data format standardisation as well as better management of personal data. Current machine learning technologies require curated and accurate data to enable companies, research institutes and the public sector to create innovative products and services. For example, participants credited the tremendous innovation taking place in the satellite sector to the open data policies for satellite data of public entities like NASA or Copernicus.

Applications of AI in earth observation and science were an important focus of the discussion. Innovative geospatial tools use machine learning technologies to derive strategic intelligence from satellite data by precisely monitoring sectors ranging from finance, agriculture and land use to disaster management, oil and gas and transportation. They also offer new opportunities to quickly map and predict economic activities of countries at micro and macro levels. The consensus was that this growing demand requires strengthening data distribution platforms and

building a data exploitation system based on advanced information and communications technologies.

AI is also permeating science, which is becoming computational. Traditional science first builds a hypothetic model based on human intelligence and then evaluates the model with data. In contrast, computational science uses "data as the model" and replaces scientific understanding with scientific exploration. AI scientists said AI systems can detect patterns and make discoveries in data that humans cannot detect, as well as explore realms of hypotheses that human cognition cannot imagine. Robot scientists conduct and optimise high-precision experiments in a documented reproducible way. Human scientists are needed to conceptualise problems and to provide feedback for algorithms but the consensus was that research institutions, particularly in bio-medical science and life science fields, will require capable AI systems to remain competitive.

Threats to individual privacy and democratic principles must be addressed

The exponential growth of AI, and corresponding consumption and analysis of big data, has underscored the need for new policies and standards to protect individual data and safeguard democratic institutions, according to participants. Recent controversies about the role of social networking platforms in misuse of personal data and allegations of interference in democratic processes illustrate the growing challenges to existing privacy frameworks. Participants discussed the role of the European Union's new General Data Protection Regulation (GDPR) as a means of strengthening data protection in Europe and stiffening penalties for its misuse. Multi-disciplinary teams were proposed as a means of embedding privacy into AI solutions and conducting privacy impact assessments to balance privacy against functionality and flexibility of the technology.

A call for fairness and accountability

A key conference takeaway was a call for transparency in AI decision-making to ensure fairness and accountability. This was particularly critical for AI-powered decisions that have an impact on individual lives. One participant said transparency in AI, often referred to as "explainability," should allow people to understand both how an AI system operates and the chain of reasoning leading to a decision. Rather than opening the "black box" of algorithms, which would require a level of technical understanding beyond most people, explainability would help users understand how AI systems are developed, trained and deployed.

Participants also emphasised the need for transparency and accountability for high-stakes AI applications in criminal justice, driverless vehicles, personal finance and healthcare. For instance, authorities and the public would need to understand the decision-making behind whether a driverless car faced with an accident chooses to hit a bicyclist or hurt its passenger. Another example cited was the need to explain how AI is used in deciding whether to short-list or hire a specific job applicant.

Similar issues arose around the impact of self-learning and autonomous systems on responsibility, safety and liability standards for consumer products. Presenters underlined the need for a broad re-thinking of how to assign liability and enforce safety standards for manufacturers and users. Household appliances were cited as an example of the need for evolving standards. Many appliances are increasingly autonomous, but existing safety standards focus on hardware rather than software. A business representative noted that the safety benefits of AI products should be considered along with the risks, and that safety and financial liability standards should match those for non-AI products.

Minimising or eliminating bias as AI moves more deeply into decision making traditionally controlled by people was also on the radar. Citing numerous examples of machine bias, some participants advocated the concept of "proportionality," which would require greater transparency for systems that affect human lives.

There was some discussion of artificial general intelligence (AGI), which is generally defined as machines equalling or exceeding human performance across the complete range of cognitive tasks. The uncertainty around the feasibility and timing of AGI made policy prescriptions difficult, but it was evident that it is an area with implications that merit consideration.

AI systems are deployed so broadly, and changes are occurring so quickly, that some participants suggested there was urgency to considering national and transnational guidelines that ensure interoperability, protect fundamental human rights, reduce inequality and promote the common good. While acknowledging the difficulty in predicting the impact of AI systems, participants representing all sectors agreed that a broad discussion is necessary to prepare for the economic and social effects.

The articulation of common principles for AI in society is needed

Common themes emerged from several sets of proposed guidelines and best practices discussed over the two days of the conference. Among the frameworks that helped inform the discussion were the Ethically Aligned Design principles and standards being developed by the Institute of Electrical and Electronics Engineers through its "Global Initiative for Ethical Considerations in the Design of Autonomous Systems;" the Asilomar Principles from the Future of Life Institute; the guidelines on AI R&D developed by the Japanese Ministry of Internal Affairs and Communications for researchers and developers of AI systems; the "Partnership on Artificial Intelligence to Benefit People and Society," which plans to develop high level of principles and guidelines to assist AI researchers and developers; the UNI Global Union Principles; Microsoft principles for the Partnership of the Future; the Principles by the Association for Computing Machinery (ACM); and the AI Initiative and the UK Principles of Robotics.

The discussions highlighted the opportunity for the OECD to build on the existing principles and knowledge with its partners to identify key principles for public policy and international cooperation in AI. The OECD was viewed as a strong forum to develop international guidelines underpinned by its focus on evidence and measurement and the effective involvement of multiple stakeholders and social partners in its work. This framework would promote AI research and application deployments that reflect the needs of society for privacy, security, safety, autonomy and self-determination. The framework would clarify high-level objectives and values to help guide the rapid development of AI and provide a normative statement of the key elements for a successful AI transition. Emphasis was placed on principles and rules that are sufficiently flexible, do not stifle innovation, and involve all stakeholders. Several participants stressed the urgency of creating an international public policy framework to help shape the future of AI and its implications, because of accelerating AI development.

The OECD's broader work on an integrated policy framework to help governments navigate the digital transformation across policy domains was also highlighted (the *"Going Digital"* project) and AI viewed as an important manifestation of the digital transformation with profound impact throughout societies -- including on productivity, employment, business models or public service -- and that requires coherent public policies. The OECD's related and ongoing work on Enhanced Access to Data was noted: analytical reports, a measurement agenda and possibly guidelines, aim to balance access to data to benefit society with basic tenets of privacy. There was also consensus on the need to measure the diffusion of AI systems more broadly than GDP, based on concepts like the OECD's Better Life Index.

Detailed Summary

WELCOME SESSION

Wonki Min, Chair of the OECD Committee on Digital Economy Policy (CDEP), Korea, opened the conference and thanked the Japanese Ministry of Internal Affairs and Communications (MIC) for supporting the OECD work on AI and the conference.

Garry Kasparov, Former World Chess Champion and author of "Deep Thinking" (by video) highlighted the timeliness of the OECD addressing policies for intelligent machines. He assimilated human progress to human labour being taken over by farm animals, manufacturing and calculations, to now white-collar jobs. Saying that technology disrupts industries before creating new jobs and opportunities, he recalled being one of the first knowledge workers whose job was threatened by AI when he lost a chess game against IBM supercomputer Deep Blue in 1998.

He stressed that humans having built Deep Blue, they would leverage intelligent machines as a tool to enhance human abilities. He added that powerful autonomous machines must reflect human morality and that the wealth, productivity and safety generated by AI must be shared across society. He cautioned that while AI empowers humans, it does not change their nature; they can use AI for good –education, communication, commerce – or for evil – propaganda or terrorism. He advised conference participants not to be afraid and try to slow down AI but instead to keep creating new non-routine tasks and directions that require uniquely human creativity.

Andrew Wyckoff, OECD Director for Science, Technology and Innovation, introduced the multidisciplinary project "Going Digital: Making the Transformation Work for Growth and Well-being" that engages 14 OECD policy communities –from tax to trade to transportation– and multiple stakeholders groups –business, civil society, organised labour and the technical community– to collectively identify the opportunities and address the challenges economies and societies face in an increasingly digital and data-driven world. He described AI – with its combinatorial links to the Internet of Things (IoT), "big data," and machine learning – as an important manifestation of the digital transformation. He said AI affects productivity, employment, business models and human relations, stressing the remarkable promise of AI discerning complex patterns, detecting irregularities and helping to allocate resources efficiently; triggering business and scientific innovations.

Cautioning against overpromising but also against hyping fear, he felt AI development was at a pivotal point and viewed the event as a milestone to discuss policy and institutional frameworks and values that should guide AI design and use. The speed of development of AI makes this urgent and means that the private sector and technologists have a vital role to play. The implications of AI are so widespread that the conversation must be inclusive and global.

Masahiko Tominaga, **Vice-Minister for Policy Coordination, Ministry of Internal Affairs and Communications (MIC), Japan**, articulated Japan's vision in which AI connects the physical and digital worlds for humans and drives economic development and social transformation, generating wisdom that enriches people's lives and benefits society at large. He highlighted challenges of AI; safety, cybersecurity, privacy and ethics. In line with Japan's AI technology strategy and industrialisation roadmap released in March 2018, research by the

Ministry of Internal Affairs and Communication (MIC) aims to build AI from social big data from understanding of brain activity mechanisms.

In addition, following the G7 ICT Ministers' Meeting in Japan in April 2016, MIC convened the "Conference toward AI Network Society", a multi-stakeholder expert group to investigate social, economic, ethical and legal issues related to AI. The group produced draft AI R&D Guidelines in July 2017. The Vice-Minister offered Japanese support and these guidelines to help advance discussions on AI at the international level, including at the OECD. He recalled that G7 ICT and Industry Ministers Meeting in Turin, Italy in September 2017 shared a vision of "human centric" AI and asked the OECD to support cooperation and multi-stakeholder dialogue on AI.

SESSION 1: THE STATE OF AI RESEARCH

Kenneth Cukier, **Senior Editor, The Economist, United Kingdom**, likened the current golden age of AI to the period 20-25 years ago with the World Wide Web and introduced the session, to discuss AI capabilities and the state of play today and tomorrow.

Francesca Rossi, **Research Scientist, IBM Watson and Professor of Computer Science, University of Padova, Italy** *(Human-AI collaboration: technical and ethical challenges)* explained that IBM focuses on AI that augments human intelligence and enables better decision making. She underlined remaining challenges to allow effective human /machine communication; proactive decision support by machines to humans; and dynamic improvement of machines over time through interactions and with less data. She also noted ethical challenges: aligning values without a universal set of values; mutual trust and transparency of AI systems and of data handling; avoiding biases; and developing workforce skills; as well as engaging all stakeholders.

Stuart Russell, **Professor of Computer Science, University of California, Berkeley, United States** *(Human-compatible artificial intelligence)* noted that AI had progressed much faster than expected. He argued that AI will make better decisions than humans with more information and foresight and is likely to replace human labour. Since new occupations such as data scientists will not provide many jobs, he predicted humans would focus on work to improve each other's lives. While such occupations are often poorly paid at present, increased professionalisation and a greater emphasis on research would lead to higher added value and incomes. This in turn suggests the need for significant changes in education and science.

He explained how to evolve AI systems design from optimising a given objective to "provably beneficial AI" for humans, based on AI techniques for learning human preferences. He explained why humans' computational limitations, inconsistent preferences, and nastiness made this task difficult. He was optimistic, however, that we could get it right as we have strong economic incentives and significant amounts of data about human choices in the written record.

Rodolphe Gelin, **Robotics Software Engineering Lead, SoftBank Robotics, Paris** *(Robots, man's best friend)* said that robots would assist humans rather than replacing them. He introduced Softbank's approach to robotics, which focuses on the business-to-business market and on using robots in education as a complement to teachers, in providing assistance to elderly people, and in future as a companion for the family. He emphasised that robot owners would prefer robots that obey them rather than ethical robots.

Osamu Sudoh, **Professor, University of Tokyo Interfaculty Initiative in Information Studies, Japan** *(Towards AI network society - addressing social, economic, ethical and legal*

issues) introduced the "Artificial Intelligence Technology Strategy Council" established by the Japanese government and its AI industrialisation roadmap. By 2020, Japan aims to have standardised and interconnected public data, to enable advanced public services to society. By 2025-35, Japan will connect various AIs to constitute society's enabling ecosystem. Japan's Council of Science, Technology and Innovation "*Society 5.0*" vision will seamlessly interconnect cyber and physical worlds.

Recalling the work by the Japanese MIC, he brought up: *i) interoperability* – AI systems are currently different from company to company, but will need to communicate to create value – for example, autonomous vehicles from different manufacturers must communicate with each other, creating the need for data standardisation; *ii) human-AI interaction* in which people co-evolve with AI and do not need to fear superintelligence, in line with Japan's proposal at the G7 ICT ministerial meeting in Takamatsu to develop AI R&D guidelines; and *iii) adjustments to address societal impacts* of AI, notably examining more sustainable social security systems for a world with robots and AI – for example, experimentation with a basic income in Finland – and evolving education systems so that people can learn how to use AI as a tool to tackle global challenges.

Philipp Slusallek, **Scientific Director at DFKI, Germany**, focused on how to ensure AI systems interact safely and reliably with the very complex real world *(Artificial intelligence and digital reality: Do we need a "CERN for AI"?)*. He defined AI as "systems that are able to perceive, learn, communicate, reason, plan and simulate in a virtual world and act in the real world", *i.e.* AI simultaneously understands the real world by learning models or the "rules of the game" and finds the best strategies to act given these models of reality. He explained that machine learning is the best method for such systems to adapt to and learn about complex interactions with the real world. However, it typically requires large data inputs for training and does not enable formal validation of results. For example, because the system cannot be trained and tested in the real world, there is no guarantee that an autonomous car will react correctly if faced with the risk of hitting a child.

He put forward the idea of a "digital reality"; a simulated environment replicating the relevant features of the real world to: *i)* generate input data to train AI systems on complex situations by modelling the real world so as to generate synthetic data, *ii)* benchmark by reproducing and standardising test scenarios, *iii)* validate performance and recalibrate synthetic data versus real data, identify and adapt incomplete models or set-up tests (like a "driver's licence test" for autonomous vehicles); and, *iv)* explore the system's decision-making process and the potential outcomes of alternative decisions.

Through this type of simulation, AlphaGo Zero was able to learn to play Go superhumanly with no data, simply by playing Go against itself. But because the "rules of the game" of reality are much more complex than Go and are not articulated, partial models based on real world data would be needed to create a digital reality. For example for driving, different models *e.g.* of pedestrian motion, traffic environments, etc. would be combined to describe scenarios, including critical scenarios for which there is otherwise no data, such as children running in front of a car (different child sizes, clothing, directions, speed, lighting conditions, etc.). The simulations would generate the required synthetic sensor data (for cameras, radar, lidar, etc.) that a real car would observe in reality in this kind of situation.

Mr. Slussalek called for common approaches to challenges and a joint research platform and community or "CERN for AI" as a collaborative scientific effort to continuously improve and understand the real world; provide a transparent, open and flexible platform supporting research and facilitating transfer/exchange with industry, and use the platform for broad discussion of policies and consequences of using AI.

The discussion focused on transparency or "explainability". Participants emphasised the need for explanations and for a human-machine interrogation process, particularly in high stake recommendations or decisions in areas such as criminal justice, personal finance and healthcare. They also noted that in many cases, there is a trade-off between explainability and accuracy/performance. AI experts explained the different technical approaches that can be taken, of either building explainability into the training or the system as such or treating the result as a scientific problem and obtaining some explainability indirectly. They said that the level of uncertainty of the AI system should be communicated to, and accepted by, the human decision-makers, although it would not be exact. Sensitivity analysis of critical variables could be provided in some cases to explain why a path of action is optimal, with multilayer explanations and decision trees. The permissible error rate is likely to vary depending on the application. For example, the 87% success rate of the Japanese Voicetra system may be permissible for translation but may not be for autonomous driving or medical exams.

Speakers also highlighted the importance of not making machines that can choose to kill human beings autonomously and the need for someone to be responsible for robots' actions. They also said that AI in its current state is not revolutionary but AI would become revolutionary when individual AI systems can understand each other, work together without humans teaching them, and act with "common sense".

SESSION 2: AI APPLICATIONS AND CASE STUDIES

Andrew Wyckoff, OECD Director for Science, Technology and Innovation, introduced the session on how AI is being applied in real life to make better decisions, reduce costs and improve productivity in a variety of domains in real life applications in health, transportation, security, as well as more unexpected areas such as arts and culture and services.

Valerio Dilda, Partner, Paris, McKinsey & Company, introduced the findings of a recent McKinsey report *(AI: perspectives and opportunities)*. In areas such as image recognition, AI performance has made tremendous process recently to now exceed humans. Since 2010 image recognition accuracy grew from less than 70% to 96% while human performance accuracy level is at 95%. He quoted Andrew Ng in saying that "AI is the new electricity" and already has serious business impact. For example, Netflix attributes USD 1 billion of revenues to prevented churn thanks to AI and Amazon credits AI with a 75% reduction on click-to-ship time. AI automatises tasks rather than jobs. 40% of tasks can be replaced but less than 10% of jobs will be replaced.

McKinsey found adoption at scale to be limited. Businesses using AI tend to do so in their core business, to boost revenues and optimise costs by identifying issues that AI can help with. By sector, digitally mature industries that have data they can leverage – finance, tech&telcos and automotive – lead in terms of current and foreseen AI investments in AI. AI early adopters also tend to be larger businesses. McKinsey estimates that AI adoption in retail can increase sales by 30%, optimise production time by 30% in manufacturing, reduce costs by 10% in electricity and lead to potential savings of over 10% in health care.

Reinhard Stolle, **Department of Artificial Intelligence at BMW AG, Munich** *(AI as a driver of the automotive industry)* described the development of autonomous driving applications by BMW. He described five levels of automation in driving whereby levels 0 to 2 are the current state of the art, level 3 involves autonomous driving on motorways using humans as fall-back, and levels 4 and 5 have no human intervention. AI, machine learning and big data are on-board vehicles, with sensors, scene understanding, action planning and

execution as well as off-board, with loop of collection data, evolution through training, and simulations.

Max Yuan, **founder and chairman, Xiaoi Robot Technology, Shanghai** *(AI empowers government and enterprises)* provided an overview of Xiaoi's conversational AI applications that process 3 million conversations daily in People's Republic of China (hereafter 'China'), reducing the number of calls by 2 million monthly and decreasing costs by about USD 14 million. He highlighted applications of conversational AI technology in: *i)* smart customer service (AI + online channels), including training and marketing; *ii)* smart cities (brain of the smart city) and smart public services (e.g. the auto-flow intelligent service system for the city of Guiyang); *iii)* smart robot applications that combine robotics and AI; *iv)* smart devices with a machine-mind operating system used in smart homes, smart vehicles, smart wearables, smart robots; and, *v)* smart offices.

Lynette Webb, **Senior Manager, European Policy Strategy, Google, London** *(Machine learning in action)* emphasised that machine learning is not magic but mathematics**,** pattern matching and probability. She explained that machine learning is an increasingly key component of Google's products. However, it requires curated and accurate data; enough computer power and tools for training; and people for set-up and results assessment. Google uses statistical machine translation to translate over a billion words daily; speech recognition that is reaching near-human accuracy levels; and image recognition to automatically categorise objects and concepts. She gave examples of applications. AIRBUS uses AI tools to automate the detection and correction of satellite imagery. A Japanese car auction service uses AI to automatically classify photos uploaded by used-car dealers. Rolls Royce detects, classifies and tracks objects that ships encounter at sea. Connecterra tracks cow behaviour via sensors and uses machine learning to warn farmers when a cow might be ill.

She highlighted the importance of openness and common norms to the success of machine learning. She explained that Google shares curated training datasets, tools and training material. She also introduced Google's recent People + AI Research initiative (PAIR) initiative across Google and added that Google participates in the Digital Ethics Lab at Oxford Internet Institute, and in the Partnership on AI.

In discussion, participants highlighted:

- The possible impact of AI on jobs and the likelihood that, as mentioned in a recent ITF report on driverless trucks, levels 4 and 5 of vehicle automation could reduce costs of trucking so much – around 30% – that non-adopters would be rapidly driven out of business and small players may have to consolidate or disappear.
- The role of AI in improving public services and in improving the efficiency and accuracy of public policy decisions.
- The respective roles of government, academia, and private actors in making available and maintaining AI technologies and data in the commons to encourage innovation:
  - While success comes with investment and trial and error, available data shows that firms investing in and gaining from AI are large firms.
  - In this context, how to enable SMEs to adopt and benefit from AI, such as through upskilling and helping SMEs to make the required investments in technology and data.
  - The role for public policy makers to focus on selected vertical industry sectors that AI could benefit. For example in France in agriculture; policies to encourage investment in specific AI applications could benefit all players while

individual players could not make these investments alone in a fragmented sector with many SMEs.

- The governance of decision-making using assisted intelligence to inform and support faster and more accurate decisions by human beings.
- Ethics by design and how building ethics into AI systems requires dialogue because the rules that guide humans and for instance, human drivers, are not written and cars will need to break some written rules.

SESSION 3: CLOSE-UP ON AI IN SPACE APPLICATIONS

Claire Jolly, **Head of the OECD Space Forum**, introduced the session organised by the OECD Space Forum that provided an overview of innovative geospatial tools that use machine learning technologies to derive strategic intelligence from satellite data and make predictions.

Tugdual Ceillier, **Lead Data Scientist, EarthCube, Toulouse** *(Artificial intelligence and remote sensing: new capabilities to monitor infrastructure)* explained how Earthcube combines satellite imagery and AI to investigate threats to pipelines. Threats to the system come from human activity in proximity of the pipelines, including, for example, deforestation, vehicles in movement, levelling works and new sites and road construction. AI helps to detect threats by recognising patterns in satellite images.

Bryan Yates, **Director of Sales - EMEA region, Orbital Insight, Mountain View, California** *(New geoanalytics: tracking economies from space)* explained how Orbital Insight applies machine/deep learning algorithms and computer-vision to process, classify, and analyse satellite, aerial or unmanned aerial vehicle (UAV) images to count, classify and detect changes across domains *e.g.* floating roof oil tank levels, water levels, crop yields and health, land and building classifications, as well as road networks, counting cars and trucks, monitoring mining activities and commodity stockpiles. Satellite imagery is analysed with machine learning and combined with financial and manufacturing data to generate insights that can increase revenue and save costs.

Thanh-Long Huynh, **CEO, Quantcube Technology, Paris** *(Big data analytics for strategic intelligence)* said that Quantcube looks at alternative sources of information to perform short, middle and long term strategic intelligence. Sources used include satellite data, social networks, professional networks, online retailers, blogs, e-commerce, meteorological data and air, marine and road traffic data. Collected data range from pictures to text and used to make predictions on energy trends (*e.g.* based on OPEC meeting news), manufacturing (*e.g.* based on images of new factories and new infrastructure, or resources (*e.g.* based on geo-localisation of ships transporting iron, ore and energy).

Bahaa Alhaddad, **Space Business Development, Starlab Space, Harwell Oxford, United Kingdom** *(Neurosciences and space data: a new big bang)* said that Starlab uses machine learning to analyse geospatial data for urban green management to monitor the health of trees within cities and to develop "happier cities" in combination with neuroscience monitoring of brain activity under different urban scenarios (presence of green areas, parks, museums, historical sites…). Localising green areas on satellite maps helps identify "happy areas" in urban agglomerations, used to suggest itineraries or to provide information on the quality of urban neighbourhoods.

Alexander Cooke, **Counsellor, Department of Industry, Innovation and Science, Australia** *(Digital Earth Australia)* presented the Digital Earth Australia (DEA) project established by the Australian government to increase the efficiency and effectiveness of public programmes and policies that need accurate and timely spatial information on the health and

productivity of Australia's landscape. DEA uses Sentinel 2 data, which are collected and analysed with machine learning techniques to model basin plans and monitor and model environmental watering, ocean, coasts and to provide dynamic ecological responses.

Christophe Roeland, **Policy Officer, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, European Commission, Brussels** *(EC perspectives on the earth observation revolution)* provided an overview of Copernicus, the European Programme for the establishment of a European capacity for Earth Observation. Created in the late 1990s, it collects and processes satellite data from Sentinel satellites, as well as sensor data from private companies and international partners. The programme provides services based on the data in the atmosphere, marine, land, climate change, emergency management, and security areas. Except in the security area, the satellite data and service components can be used by citizens, authorities, researchers and companies on an open and free basis. He also underlined challenges related to the huge scale of data that the Copernicus programme generates and described the European Commission project to develop a data cloud based environment – the Data Access and Information Service (DIAS) – to give users access to data and to data processing capability in the cloud where they can run their algorithms.

Some specific *space-related takeaways* follow:

Entrepreneurial companies in many parts of the world are leveraging machine learning and a "deluge" of satellite data to build innovation products and services. The key role of public policies was underscored, in particular: *i)* the importance of open data policies for publicly-funded satellite data, to enable SMEs to create value-added products and services leveraging, for example, the Copernicus fleet of satellites developed by the European Space Agency and funded by the European Union or by Landsat in the United States; *ii)* the need for interoperability and standards notably for data archiving and data consultation; and, *iii)* future skills needs in view of the competition for scarce machine learning scientists and engineers.

SESSION 4: ENHANCING DISCOVERY – THE ROLE OF AI IN SCIENCE

Dominique Guellec, **Head of OECD Science and Technology Policy Division,** introduced the session on AI in science, of which the goal was to explore opportunities and challenges of applying AI in science. Discussion items included current and emerging uses of AI and machine learning in science; limitations in using AI in science; the opportunities AI offers to increase research productivity; challenges posed for researchers; and, issues raised for science education, research sponsors and policymakers.

Stephen Roberts, **Professor of Machine Learning in Information Engineering, University of Oxford, United Kingdom** *(21st century science: the age of intelligent algorithms)* highlighted the driving forces of AI developments; the growing scale of data generated by people, progress in machine learning algorithms and exponential growth of affordable computational capability. He emphasised the less visible but very important role of AI in curating large volumes of data. He stressed the new scientific paradigm driven by data science in which data acquisition is the priority and human scientists augment discoveries made by algorithmic science in large experiments.

He distinguished the main applications of AI and machine learning in science as: *i)* Detection and discovery: for example, analysis of large scale data enabled detection of planets in the Kepler space telescope project; *ii)* Deriving principles from data observation: for example, Newton's law of motion could be deduced from observing data on falling objects; *iii)* Human-in-the-loop, *i.e.* combining human wisdom and knowledge with algorithms to create effective feedback mechanism to solve global challenges; and *iv)* Smart experiments, whereby AI

optimises experimental design for large-scale experiments such as designing new materials, new drugs, or new treatments. He concluded that the era of automated science has begun.

Hiroaki Kitano, **President and CEO of Sony Computer Science Laboratories, Japan** *(The Nobel Turing Challenge: creating the engine of scientific discovery)* introduced a proposed "Nobel Turing Challenge". The goal is to incite the development of AI systems that can make major scientific discoveries meriting a Nobel Prize and that can execute scientific activities autonomously, such as choosing a topic and communicating with the community, like a human researcher. The primary focus of the challenge is on biomedical science for the Physiology and Medicine Award.

He said scientific discovery was limited by human cognitive capacity and reliance on researchers' serendipity or intuition. AI could help science to conduct massive search and verify entire hypotheses space with increasing data and massive computing power. He identified three strengths of AI for scientific research: *i)* detecting patterns that human cannot detect, *ii)* performing high-precision experiments, and *iii)* exploring hypotheses that humans cannot imagine. He demonstrated Beatles-style music that had been fully composed by AI and was hopeful that knowledge created by AI could make a qualitative difference in the evolution of human civilisation.

Ross King, **Professor of Machine Intelligence, Manchester University School of Computer Science, United Kingdom** *(The automation of science)* identified how AI systems augment scientists with flawless logical reasoning, learning from vast amounts of data, and reading millions of scientific papers. He introduced Robot Scientists that combine AI systems and laboratory automation to originate experiments, physically execute them, interpret the results, and repeat the experiments cycle. He said that Robot Scientists could work harder, faster, more efficiently, were easier to reproduce (than human scientists that must each be trained from scratch), and were helping to improve knowledge and share data. He illustrated his talk with a video of the robot scientists called "Adam" and "Eve". He described the work of these robots on drugs against tropical diseases – they had found that a compound called "Triclosan Repositioned", commonly used in toothpaste, is effective against Malaria. He said that just like teams combining both humans and computers play better chess than either alone can play, human and robot scientists could collaborate to improve the productivity of science for societal benefits such as better food security and better medicines.

Jonathan McLoone, Technical Director, Wolfram Research Europe *(Preparing science for AI: rethinking education, research and publication)* described how AI and computational thinking require rethinking practices in *i)* scientific discovery, *ii)* dissemination of science, and *iii)* scientific education. Regarding discovery, he underlined that computation thinking is a process of exploration rather than one of understanding. While traditional science first builds a hypothetic model based on human intelligence and then evaluates the model with data, computation science uses data as the model. Another shift is from a simplicity mind-set toward a complexity mind-set: traditional science aims to eliminate noise to evaluate model through experimentation, while computational thinking require as much data as possible, including noise that could help signal the solution.

Traditional scientific dissemination and publication processes rely on peer review, whereas reproducibility is critical in computational thinking. Reproducibility requires providing access to machine learning experiment code as well as machine-readable data. He called for reforms to incentivise the provision by scientists of code and data in scientific publications to allow reproducibility.

He stressed the importance of adapting education to prepare students for computational thinking by: adapting mathematics education to prioritise conceptual work instead of hand written calculation (in current mathematics classes most of the time is spent on simple hand written calculation which are too simple for real scientific issues). He recommended that policy makers adapt the traditional system of education and scientific dissemination to prepare for computational thinking and machine learning experiments in science.

In discussion, participants raised issues which may need policy attention:

- Governments control public health data assets that could be leveraged with AI for scientific discoveries, but participants felt that current policies favour privacy over scientific progress.
- Current AI research and development is being led by private company investment, but governments and public research play a key role in leveraging AI to help solve global challenges.
- In the near future, research institutions will require capable AI systems to remain competitive, particularly in bio-medical science and life science fields.
- Scientific discovery powered by AI requires large scale capital investment to prepare rich data resources, automated laboratories and massive computing facilities. This may lead to the concentration of scientific discovery and raises concerns about excessive monopoly of scientific knowledge.

SESSION 5: THE AI POLICY LANDSCAPE

Anne Carblanc, Head of OECD Division on Digital Economy Policy, introduced session 5, the goal of which was to provide an overview of the AI policy landscape, covering initiatives by governments, the private sector, research communities, civil society and trade unions. She underscored that fostering the beneficial development of AI had become a priority not just at the national level – in Japan, China and Finland for example – but also at the international level, such as at the G7 and EU levels.

David Heiner, Strategic Policy Advisor at Microsoft, representative of the Partnership on Artificial Intelligence *(Enabling the promise of artificial intelligence)* noted that Microsoft and other AI firms recognise that issues such as fairness, transparency and accountability must be addressed so that people can trust AI. He said that AI excels at narrow tasks and can reason over vast amounts of data, turning video, speech, and sounds into text with image recognition. But extracting meaning from this text is difficult and requires people with human qualities such as empathy, judgement and fairness. AI can allow people to interact with computers more naturally with improved, natural, interfaces and can be designed to amplify human ingenuity.

He emphasised that AI should not be controlled by a small number of US companies and noted that Microsoft was making its image recognition AI technology broadly available, building end user services and exposing the programming interfaces to developers so that anyone can develop new applications. He presented Microsoft's framework to amplify human ingenuity to benefit everyone and pointed out that AI could also help address each issue, *e.g.* self-driving cars improve reliability and safety; computers may not have the same biases as humans; AI agents can help improve security and privacy. He introduced the "Partnership on AI to benefit people and society", which held its first meeting in Berlin the same week discussing fairness, transparency and accountability and planned to develop high level of principles and guidelines to assist AI researchers and developers.

Nicolas Miailhe, Co-Founder & President, The Future Society; Senior Visiting Research Fellow, Harvard Kennedy School of Government *(Harnessing the power of collective*

intelligence to govern the rise of AI: the case of "algorithmic transparency & accountability") introduced the "Global Civic Debate on the Governance of AI" he is leading. In his view AI issues are big data issues, since AI is fuelled by growing stocks and flows of data. In his view a key question is therefore how the private and public sector monetise data and share the value with consumers. He viewed the European GDPR as key to data portability.

Miailhe advocated for a "regulatory sandbox" for AI to allow actors to play with data at scale away from legal constraints because: *i)* algorithmic correlation weakens the distinction between personal data and other data, and non-personal data could allow the reidentification of an individual; *ii)* AI weakens the distinction between experimentation and deployment of AI; *iii)* countries seeking sovereignty in the field of AI must balance competing imperatives of privacy and controllability, with free data flows used by both private and public actors. He also pointed out the competing objectives of dignity/privacy versus access to the technology.

Benedetta Arese Lucini (Italy) *(G7 Italy: towards a human-centric AI)* presented the result of the G7 ICT and Industry Ministerial held in Turin in September 2017 under the Italian presidency. The G7 ICT and Industry Ministerial Declaration was the first internationally-agreed text about AI. The G7 countries acknowledged the tremendous potential benefits of AI but also its uncertain impact on society and economy. They agreed to take a "human centric" approach to AI and to *i)* gain understanding of the cultural, ethical, regulatory, and legal impact of AI, *ii)* note the outcome of multi-stakeholder dialogue during the Ministerial to explore both the positive and controversial impact of AI, notably on growth, job creation, accountability, privacy and security, *iii)* pursue a multi-stakeholder approach, as an effective way to address policy and regulatory issues, and *iv)* work towards common understanding of how to benefit from the full potential of AI for equitable society, while underlining that regulation must not hinder the development of technology and industry.

Susumu Hirano (Japan), Faculty of Policy Studies / Professor, Dean, Graduate School of Policy Studies, Chuo University *(AI R&D guidelines)* recalled the April 2016 G7 ICT Ministerial Meeting of Takamatsu (Japan). At that time Ms Sanae Takaichi, the Japanese Minister of Internal Affairs and Communications, proposed that G7 countries lead international discussions on a non-binding international framework for AI development, building on the set of AI R&D Principles developed by her Ministry. The Japanese Ministry of Internal Affairs and Communications (MIC) had then created a "Conference toward AI Network Society" in October 2016, to develop proposed "AI R&D Guidelines".

He introduced the guidelines of July 2017 that provide 9 principles that researchers and developers of AI systems should follow.

- In order to promote the benefits of AI systems: *i)* collaboration;
- In order to mitigate the risks of AI systems: *ii)* transparency, *iii)* controllability, *iv)* safety, *v)* security, *vi)* privacy, and *vii)* ethics;
- To encourage acceptance by users and other stakeholders: *viii)* user assistance and *ix)* accountability.

Their overall purpose is to achieve a human-centered society, balance benefits & risks of AI networks, ensure technological neutrality and avoid excessive burden on developers. To this end, MIC aims for the Guidelines to become non-binding soft law and best practice among stakeholders internationally, and that they be reviewed constantly and revised flexibly as needed.

Cédric Villani (France), député LREM de l'Essonne, **chargé de mission IA** *(Overview of AI policy initiative in France)* said that he had been tasked by the President to recommend AI policies to the French government and to Europe, where GDPR enforcement would begin in

2018. The first goal of the mission is to enhance the attractiveness and performance of France and Europe; the second is to consider justice and social linkages. He said that the situation needed to be better balanced and highlighted concerns about the flow of researchers and engineers from Europe to other regions and from the public sector to the private sector.

Mr. Villani shared preliminary conclusions that:

i. AI is transversal, requiring broad whole-of-society involvement in a public debate on AI;
ii. Current AI technologies rely on data use. Data protection, data regulation, and the data economy are complex issues that are legal, cultural, and technical. Companies engaged in data collection need to respect Europe's will to protect privacy at every stage of the process;
iii. The need for Europe to be more competitive and networked to attract scarce human resources – notably economists, mathematicians and statisticians – who can develop, programme and use AI innovatively.

In closing he noted the widely varying predictions by economists and that while AI was simply making correlations, he believed that the future would involve models and causality.

Xiao Zhang (China), Vice Director, China Internet Network Information Center *(Overview of China's digital economy and AI policy)* explained that following the Internet applications phase and the mobile Internet phase, China is currently in the third phase of digitalisation that is being driven by the Internet of Things (IoT), with AI expected to integrate future data and systems. She explained that digitalisation efforts in China focus on strengthening: *i)* access connections to mobile and IoT; *ii)* platform ecosystems, *iii)* big data; and *iv)* intelligence and AI R&D. She introduced China's "Three-year Guidance for Internet Plus Artificial Intelligence Plan (2016-2018)" that focuses on: *i)* enhancing AI hardware capacity, *ii)* strong platform ecosystems, *iii)* AI applications in important socioeconomic areas, and *iv)* AI's impact on society. She also introduced the "Guideline on AI Development" which provides China's long-term perspective on AI and aims for China to be a global AI innovation centre by 2030.

Pekka Sivonen (Finland), Director, Digitalisation Strategy and Programmes, Tekes Innovation Funding Agency, introduced Finland's vision of the autonomous society in the coming 10 years *(Ambitious development programs enabling rapid growth of AI and platform economy in Finland)*. He presented Finland's national strategies toward the development of a platform economy and for AI: Finland aims to develop a safe and democratic society with AI, to provide the best public services in the world and for AI to bring new prosperity, growth, and productivity to citizens. He explained that Finland aims to take a holistic approach to AI and to the platform economy where each part of the value chain interacts with each other; driven by the internet, blockchain, AI, and new display technology. Finland has some 250 companies working on AI development.

Marten Kaevats (Estonia), National Digital Advisor, Government Office of Estonia, gave an overview of AI deployment in Estonia *(Estonia's ideas on legalising AI)*. He said that Estonia is constantly testing new e-government ideas to save costs and improve efficiency, for example basing all governmental information systems in Estonia on blockchain as early as 2012. Estonia is planning the next step of its e-governance system that is powered by AI and experimenting with e-healthcare and situational awareness. The focus of the Estonian discussion is on improving lives and cities and supporting human values, with a focus on ethics and liability. On the enforcement side, Estonia focuses on core values of ethics, liability,

integrity, and accountability and is building an enforcement system based on blockchain that mitigates integrity and accountability risks, with a pilot project planned in 2018.

With "StreetLEGAL" self-driving cars can be tested on Estonian public roads since March 2017. Estonia is also the first and only government discussing the legalisation of the AI; *i.e.* giving representative rights, and responsibilities, to algorithms to buy and sell services on their owners' behalf. The government is considering several options and aims to have a bill ready by late 2018.

Cécile Huet (European Commission), **Deputy Head of Unit, Robotics and Artificial Intelligence, DG CONNECT, Brussels**, explained *the European Commission's initiatives on AI* aiming to support AI development to increase efficiency and flexibility, facilitate interaction and cooperation, enhance productivity, competitiveness and growth, and improve quality of citizens' life at work and home. She said that the European Commission was trying to overcome barriers to AI deployment related to: *i)* technological performance issues, by funding EU robotics and AI programmes within the "Horizon 2020" project (EUR 700 million for robotics, EUR 20 million for an IoT pilot project on autonomous vehicles and EUR 20 million to develop an AI-on-demand platform); *ii)* ethical, legal and societal issues related to safety, liability, data protection & ownership, and employment – the Commission is evaluating existing Directives on Machinery and Products Liability for AI-related safety and liability and will support standardisation efforts through test bed experimentation; and *iii)* public acceptance, with for example the "European Robotics Week" opening laboratories to the public. She welcomed further co-operation with the OECD on common issues.

The discussion focused on identifying commonalities and differences between and on the role of policies to benefit from AI while addressing its challenges. In the policy approaches towards AI systems:

- All the strategies and initiatives presented all aim to foster growth, productivity and competitiveness.
- There was broad agreement that AI systems should be "human-centric" and that collective and societal decisions are needed on the values and choices that should guide AI systems. There was also agreement on the need to developing principles to help developers and researchers build systems that are safe and fair.
- Differing legal systems, cultures and political contexts play a critical role, because AI systems are integrated into existing industries, healthcare systems, education systems etc.
- Data access and ownership for AI is a critical and sensitive issue.
    - As GDPR is rolled out across Europe in 2018, the sensitivity of personal data issues and private life in Europe is heightening caution and public debate.
    - Countries such as Finland and Estonia are moving forward fast. Estonia considers that citizens own their data while governments and companies merely provide the service of keeping the data. The Finnish strategy considers that citizens should own the data collected through their equipment. More generally Nordic countries aim to leverage the "mydata.org" structure and policies and to build a mechanism for people to benefit from and monetise their data safely.
    - "Data commons" governance models could be leveraged in Europe to help manage public, private and other types of data flows.

Participants also highlighted:

- The tension between on the one hand, scale in platform businesses and on the other hand, effective competition and lowering barriers to access.
- The roles of different stakeholders and of international co-operation:
  - The key role of governments and intergovernmental organisations in areas such as setting or codifying rules of the game, in respect to societal and economic impacts, planning for the long-term and international co-operation.
  - The convening role of governments in bringing together all stakeholders and in ensuring that AI development is inclusive.
  - International co-operation at a regional level *e.g.* among the Nordic countries, at the European level notably with the EC and at the international level, *e.g.* with the OECD. Coordination among different organisations and with China was also mentioned.
  - The need for collaboration for the future development of AI, *e.g.* to develop a "CERN for AI".
  - Participating in discussions on using AI to address global issues such as SDGs and climate changes.

SESSION 6: EMPLOYMENT & SKILLS

Mark Keese, Head of the OECD Division on Skills and Employability, introduced the session on jobs and skills and said that the exponential growth of the capabilities and applicability of AI should improve the efficiency with which goods and services are produced. All else equal, AI should therefore make everybody better off as productivity increases and prices fall. However, it also raises concern about job automation and the possibility of technological unemployment, as well as about its downwards impact on the wages of workers who are most at risk of being displaced. He explained: "some commentators have raised the spectre of inventing ourselves out of existence by developing ever more powerful AI. But a more immediate concern is whether we are running the risk of inventing ourselves out of work."

Stuart Elliot, Director of Board on Testing and Assessment, United States National Academy of Science, detailed his research to estimate the extent to which current technologies can answer the literacy and numeracy questions of the OECD Survey of Adult Skills (PIAAC) *(AI and the future of skill demand)*. His research suggests that only 11% of adults are currently above the level that AI is close to reproducing in terms of literacy skills. Although there are things that AI cannot yet do, he stressed that many individuals cannot do them either. In discussion, he also cautioned participants on the difficulty of designing education policies to bring 80% or even 50% of adults above the current computer level and suggested new tools and incentives for promoting adult skills or combining skills policies with other interventions, including social protection and social dialogue.

Frank Levy, Rose Professor Emeritus, Massachusetts Institute of Technology, added to these concerns *(Computers and populism)*. He argued that by focusing on what AI might do in the long-run, we risk missing what is happening in the short-run. Indeed, those in lower- to mid-skilled occupations involving significant amounts of repetition have already been affected by technology and remain those at highest risk of losing their jobs. Across many countries, labour markets have been polarising, with shares of both high- and low-skilled jobs rising while the share of medium-skilled, routine jobs, falls. Such upheaval in the labour market could cause severe political reactions and, argued Professor Levy, impact the adoption of AI and the policy response to it.

Christina Colclough, Senior Policy Advisor, UNI Global Union, echoed that nothing is inevitable in how and when AI is adopted *(Putting people and planet first: ethical AI enacted)*. She warned against "technological determinism" and urged social partners to talk about and agree on the kind of future we want. She presented UNI's top 10 principles for the future world of work, for workers' data privacy and protection, and for ethical artificial intelligence. UNI's top principles for AI are: *i)* AI systems must be transparent; *ii)* AI systems must be equipped with an "ethical black box"; *iii)* AI must serve people and planet with codes of ethics for the development, application and use of AI; *iv)* Adopt a human-in-command approach with responsible, safe and useful development of AI; *v)* Ensure a genderless, unbiased AI; *vi)* Share the benefits of AI systems; *vii)* Secure a just transition and ensure support for fundamental freedoms and rights; *viii)* Establish global governance mechanism; *ix)* Ban the attribution of responsibility to robots; and, *x)* Ban AI arms race.

Young Tae Kim, Secretary General, International Transport Forum *(ITF)* *(New transport for the new digital age)* said that "people have a tendency to focus on job destruction rather than on job creation", stressing that while AI may take over some tasks and jobs, it would also create a whole host of new ones. He underlined the opportunities in terms of safety, cost, sustainability and inclusiveness of leveraging AI systems in transportation. He also pointed to challenges of conflict between existing and new modes of transportation (licencing system, political influence, etc.), the need for institutions to evolve and ethical issues. In the driverless trucks, he pointed to potential responses proposed by the ITF Temporary permit system to manage the transition: temporarily requiring permits for driverless truck to have some control on the rate of deployment of driverless trucks; and using the revenues from permit sales to fund retraining efforts.

James Hairston, Head of Public Policy, Oculus VR, Facebook *(AI, employment, and general purpose technologies)* provided an overview of the use of AI at Facebook notably: *i)* machine learning that allows computers to learn and to solve problems from data without explicit programming; *ii)* computer vision that allows computers to understand visual content like images and videos, and recognise faces; and *iii)* Natural Language Processing (NLP) that allows computers to read and understand text**.** He categorised AI applications in the areas of: *i)* perception, *ii)* understanding and learning; *iii)* prediction, and *iv)* planning. He underlined that AI is not only a replacing technology, but also a complementing one and, as populations age, can help increase productivity and help deliver education. He also stressed that making AI work for everyone requires having the right policies in place and, in particular, investing in people and skills and facilitating labour mobility so that individuals can seize opportunities.

SESSION 7: PRIVACY & SECURITY

Katarina de Brisis, **Deputy Director General at Ministry of Local Government and Modernisation, Norway, Chair of OECD Working Party on Security and Privacy in the Digital Economy,** underlined the dual-use nature of AI that can both help fight cybercrime and cause harm when developed and applied with malicious intent. She cautioned about an arms race between governments, enterprises and ill-intentioned agents. She stressed that AI feeds on data, much of which is personal data that can be aggregated and de-identified but that provides significant knowledge on individuals.

Peter Fleischer, **Global Privacy Counsel, Google** *(Privacy and AI: designing machine learning systems to respect privacy)* introduced AI as requiring: *i)* computational resources, *ii)* training data, *iii)* algorithms and tools, and *iv)* creativity and ingenuity. He said that although human creativity and ingenuity progress slower than the other factors, they can achieve astonishing progress with computational resources, training data and tools that are growing

more powerful algorithmically. He emphasised that current AI is able to understand a complex photo such as that of a marathon but also people's emotions, with significant implications for privacy. Recalling the lasting importance of the OECD privacy framework, he argued that it applies in the age of AI, but that attention and scrutiny is needed on: *i)* profiling and what kind of profiling is acceptable, done by whom; *ii)* automated decision-making and which types of decisions could rely on machines only; and *iii)* spotting and correcting algorithmic bias that often come from the training data.

Taylor Owen, **Assistant Professor of Digital Media and Global Affairs, University of British Columbia** *(Governing digital infrastructure)* cautioned participants that AI facilitates misinformation on social media platforms that rely on AI (*e.g.* newsfeed algorithms) and on large-scale data collection and monetisation. He said that platform and third-party data brokers use detailed user profiles and users' inferred moods and desires to customise goods and content for them for profit, unpredictably and with biases. In his view algorithms' automatic micro-targeting combined with personal data profiles fragment the collective conversation and facilitate the proliferation and monetarisation of misinformation. He flagged governance challenges: *i)* public spaces are governed by private corporations whose interests may or may not align with the public interest, and *ii)* governments are ill-equipped to regulate large, complex and rapidly evolving platforms, such as the ads of election candidates running numerous simultaneous micro-targeted advertisements. He called for governance and oversight of algorithms to combat misinformation.

Mathias Cellarius, **Data Protection and Privacy Officer, SAP** *(AI: challenges and opportunities for data protection)* highlighted the misalignment between competing goals of *i)* effectively benefitting from data-driven AI technologies and business models; while *ii)* abiding by traditional data protection principles on purpose specification, data minimisation and use limitation. He underlined that AI requires access to large amounts of high-quality data but much of this data could be categorised as personal data under the EU's new General Data Protection Regulation (GDPR) expanded scope, placing companies and authorities in a difficult position.

He put forward that Europe should focus on safeguarding individuals' right to determine the governance of their personal information and advised against implementing the GDPR with a "one size fits all" approach, since eliminating every remote privacy risk could jeopardise valuable data uses in return for small privacy gains. He noted that approaches foreseen under GDPR – notably pseudonymisation, privacy impact assessments and privacy-by-design – could reduce the impact of personal data use on privacy and allow data utility while providing meaningful controls of personal data to users.

Kenneth Cukier, **Senior Editor, The Economist, United Kingdom** *(Do privacy laws obstruct beneficial uses of data?)* recalled that the OECD Privacy Guidelines were first issued in 1980 – a different era. He also said that AI and machine learning called into question the principles of collection limitation (since more data might generate more benefits), purpose specification (by not permitting new beneficial uses of data that are not known at the time of collection) and use limitation (why should collected data be destroyed after use since reusing data later with new techniques might generate benefits).

He said that current privacy rules can prevent beneficial uses of data; *e.g.* by preventing the use of electronic health records to predict who may catch an infectious disease. His proposal was to shift the regulatory focus from the collection to the use of data; to develop institutions and practices for data sharing and analysis; to accept that protections like anonymisation will be imperfect; to sanction entities that fail to process personal data for clear social benefit; and to develop "national privacy strategies".

In discussion, Marc Rotenberg highlighted the important distinction between *i)* data that has no personal content and no privacy consequence and that is aggregated and de-identified, in areas like climate change and education development; and *ii)* data that has privacy consequences when it is processed, *e.g.* data used in making credit determinations. He recalled that the core concern in data protection is the fair processing of information about individuals; so that when a determination is made individuals know its basis, the contributing factors, and can object if the determination is incorrect. He stressed the need for algorithmic transparency to establish accountability for computer-based decision-making. In criminal sentencing by AI, the basis for the determination is an opaque AI.

There was broad agreement in the discussion that people must be able to understand what happens to their data, the purpose of the processing, and that machines and algorithms should be under human individuals' control. Speakers noted that current AI technology can allow anonymised data that is de-identified to be linked back to individuals. The need for data security of personal data being collected and stored was also stressed.

SESSION 8: SAFETY, RESPONSIBILITY & LIABILITY

Wonki Min, **Chairman, OECD CDEP, Korea,** opened the session to focus on questions of safety, responsibility and liability raised by AI-driven automated decision-making such as autonomous vehicles and other IoT machines and "explainability".

Rod Freeman, **international products lawyer, Partner at Cooley, United Kingdom** *(Evolution or revolution? The future of regulation and liability for AI)* stressed that different types of AI applications will call for different types of policy, legal and regulatory responses that are flexible and adaptive and thus fit for purpose over the long term and that consider benefits of AI alongside risks. A survey conducted by the European Parliament on the future of robotics had found broad agreement on the need for a policy and regulatory response to AI. He recalled that product regulation is usually based on "Producer Responsibility" whereby the seller certifies safety. But current or foreseen AI applications question increasingly this model and society must decide on the liability regime applicable to AI.

Hans Ingels, **Head of Unit, Single Market Policy, Mutual Recognition and Surveillance, European Commission, DG GROW** *(Artificial intelligence and EU product liability law)* introduced the European regulation on product safety and liabilities. The Product Liability Directive (Directive 85/374/EEC) of 1985 establishes the principle of "liability without fault", *i.e.*, if a defective product causes damage to a consumer, the producer may be liable even without negligence or fault on their part. The European Commission is evaluating whether the Directive is still fit for purpose in the AI era. Preliminarily, manufacturers and insurers seem satisfied with it while consumer associations find the burden of proof for defectiveness too heavy. There are questions on how concepts of "product" and "defectiveness" apply to AI, on whether AI is a software product, on producers' liability and on the injured party's burden of proof. The Commission plans to publish AI policy reports in 2018 and create a package of policy considerations examining safety and liability issues holistically.

Pierre Chalançon, **Chair of the BIAC Consumer Task Force and Vice President Regulatory Affairs, Vorwerk & Co KG, Representation to the EU** *(Science-Fiction is not a sound basis for legislation)* first underlined the business community's view that discussions on product safety and liability rules for AI should focus on current or foreseeable technology developments and not science fiction. He said business was satisfied overall with European product safety rules although safety standards, for example, could be improved. In the household appliance industry, old safety standards focus on hardware rather than software and

regulate "finished products" while appliances increasingly make decisions in autonomous or semi-autonomous ways. If a product is a combination of hardware and software, a product's safety may need to be reassessed each time software is updated. Regarding product liability, business believes that the manufacturer who places finished products on the market should be liable for both hardware and software parts. He also stressed that cybersecurity could impact product safety if connected products are not sufficiently secure and for example, hackers take control of them at a distance and change settings.

Georg Borges, **Professor, Faculty of Law, Saarland University, Germany** *(Liability for machine-made decisions: gaps and potential solutions)* said that the liability system must evolve to account for machine-made decisions. In the automotive sector for example, the human driver is liable for damage by a car under a *fault-based liability regime* in traditional Tort Law. However, the mere passenger of an autonomous car cannot be at fault or have breached a duty of care. He noted that some suggest the "registered keeper" could bear the liability of autonomous cars, but that the "keeper" must be able to control the risk, which the keeper cannot do with autonomous cars.

He concluded that autonomous actions by machines are not covered by current fault-based liability schemes and that based on risk controllability, strict liability should be placed on manufacturers. He proposed that an insurance system fill the liability gap for autonomous systems. The European Parliament for example introduced the idea of compulsory registration of autonomous machines to create an insurance regime: based on risk assessments, registered autonomous machines would be classified, clarifying who bears the risk they pose and who is liable as a result of the insurance system. He noted challenges to develop an appropriate liability system for machine-made decisions: *i)* addressees of liability (both manufacturers and users, because both parties could contribute to the safety of autonomous machines); *ii)* liability principle (strict liability with a differentiated insurance system); and *iii)* enforceability and burden of proof.

The discussion highlighted:

- A consensus on the need for international cooperation and dialogue, particularly among regulators, to involve all stakeholders. To some extent private companies develop actual AI policies. The "co-regulation" taking place in the area of cyber security policy in Germany was provided as an example of effective governance: industry suggests security standards and if the federal agency finds them appropriate, they become a safe harbour for industry.
- The need for a holistic approach that considers safety, liability as well as the cybersecurity implications of AI.
- The importance of basing discussions on actual or foreseeable technology.
- The need to level the playing field and ensure that autonomous systems have the comparable monetary liability as *e.g.* an impaired driver.

SESSION 9: TRANSPARENCY, OVERSIGHT & ETHICS

Douglas Frantz, Deputy Secretary-General, OECD, underlined the importance, urgency and complexity of developing smart policies to ensure transparency, oversight and ethics of AI and noted that governance options range all the way from self-regulation through to government regulation or even an international treaty.

Konstantinos Karachalios, Managing Director of the IEEE-Standards Association *(The role of technical communities in making intelligent technologies work for the benefit of humanity)* introduced the Version 2 of the IEEE's practice-oriented "Ethically Aligned Design

of Autonomous and Intelligent Systems (AIS)" for comments and highlighted the duty of the technical community to self-reflect, learn from others, engage in global democratic dialogue and build bridges for collaboration at international, national and regional levels. The P-7000 series of standards seek to teach technologists to take into account contextual aspects such as ethical values into account at the systems design phase.

He stressed the challenges of first, a pace of technological evolution so rapid that political processes cannot keep up and the law cannot be adequately enforced, threatening democracy. He said secondly, that there is a misguided focus on singularity / super-intelligence when data and third-party data control models by technology giants or governments are the real, and political, threat to humanity. In his view, individuals should regain agency over their data to avoid being controlled and that the stakes for democracy, political freedom, and self-determination were very high. Third, he underlined the need to measure the impact of AIS on well-being more broadly than profit or GDP, based on concepts like the OECD's Better-Life Index.

Noting that these challenges were also opportunities, he stressed the need for political actors to reclaim their territory with the help of technologists, to ensure that AIS serve humanity – wellbeing, self-determination and the planet. All stakeholders should start to work where we can and start now. The IEEE is encouraging large-scale self-reflection within the techno-scientific community. He offered the IEEE's technical advice and assistance to policy makers and in turn sought policy makers' advice too, to help technologists understand social and political implications of technological development.

Joanna Bryson, Reader at University of Bath, and Affiliate, Center for Information Technology Policy at Princeton University *(Current and potential impacts of artificial intelligence and autonomous systems on society)* defined intelligence as doing the right thing at the right time. She said that AI is an artefact that is deliberately created by humans, for which someone is responsible and that involves computation – a physical process requiring energy, time and space. She noted that the first AI artefact in human history was writing as a way to store ideas that triggered exponential development of humans, that intelligence allowed communication and agility and the discovery of new equilibria of mutual benefits. She said that AI is blurring the distinction between customer and employee and "free" services and increasing our dependence on information bartering. It reduces the costs and advantages of geographic location, increases inequality – and thus polarisation – and transnational interdependence.

On regulating AI, she cautioned against capping liabilities and against creating incentives for more complex code, saying that AI systems' output should be regulated rather than the systems themselves. We have to cooperate to make sure things would work more for people. She agreed on the importance and complexity of issues of data control and questioned: *i)* whether a country like China can allow humans to thrive and maintain dignity while controlling peoples' lives and data; *ii)* the role of AI to help migrants with integration and language translation; *iii)* whether arts and humanities can remain interested in humans when humans can use AI to search for their own next move.

Carolyn Nguyen, Director of Technology Policy, Microsoft *(Designing AI to earn trust)* emphasised the need for wide availability of AI to people from both developed and developing countries, based on an ethical framework for trustworthy AI that includes principles of "Human-centered AI", safety, fairness, transparency, privacy and inclusiveness. She focused on fairness and transparency. She said that fairness involves treating people equally with respect and dignity, although the concept differs between countries and societal or cultural context. She put forward approaches to promote fairness/combat biases in AI: *i)* for decisions

that impact people's lives (such as a sentencing) AI-based scores should never be the sole factor while users / deciders should understand its limitations; *ii)* organisations should attract diverse types of AI talent including sociologists, economists and users, develop analytical techniques to detect bias, and *iii)* develop guidelines for developing and deploying AI systems.

Transparency or explainability in AI systems is about describing how the system is trained or procedures for systems development and deployment rather than sharing the actual code or actual data, to allow people to understand how the systems operate and to provide accountability mechanisms that ensure the development of accurate and intelligible algorithms. She noted that research is taking place on interpretability and accuracy of algorithms in groups such as Fairness, Accountability, and Transparency in Machine Learning (FATML) where AI systems are used as part of the solution. She stressed the importance of multi-stakeholder dialogue to identify and prioritise issues and mitigate risks as well as to develop and share best practices in addressing transparency in AI system deployment.

Seán Ó hÉigeartaigh, Executive Director of Cambridge's Centre for the Study of Existential Risk, introduced the *Asilomar Principles*, a set of 23 principles for the safe and socially beneficial development of AI in the near and longer term that resulted from the Future Life Institute's conference of January 2017. He said that the Asilomar conference extracted core principles from discussions, reflections and documents produced by the IEEE, academia and non-profit organisations while reflecting Western culture, particularly regarding privacy and human rights.

The issues are grouped into: *i) research issues*, with a call for research funding for beneficial AI that include difficult questions in computer science; economics, law and social studies; a constructive "science-policy link"; and a technical research culture of cooperation, trust and transparency; *ii) ethics and values*, with a call for AI systems' design and operation to be safe and secure, transparent and accountable, protective of individuals' liberty, privacy, human dignity, rights and cultural diversity, broad empowerment and shared benefits; and *iii) longer-term issues*, notably avoiding strong assumptions on the upper limits of future AI capabilities and planning carefully for the possible development of artificial general intelligence (AGI).

In discussion, participants agreed on the need to establish a policy framework to clarify various interests and objectives and to help guide the rapid development of AI, including basic benchmarking and a normative statement on what a successful AI transition should look like. They stressed the importance of working together to develop a governance framework and rules that are flexible enough, involve all stakeholders and that do not stifle innovation.

The discussion highlighted the role of different stakeholders in governance of AI:

- The complimentary governance role of governments developing relatively stable policies and of professional standards organisations such as the IEEE that can adapt and update their standards more frequently.
- Technical standards such as those of the IEEE – developed in voluntary and bottom-up processes – which were described as a non-normative governance system that is used if it brings value.
- The important role of self-regulatory efforts.
- Agreement that the OECD is well situated to develop principles for AI in society, building on the Japanese R&D Guidelines, the IEEE's Initiative, the Asilomar principles and other sets of principles such as the ACMs'.
- The need to ensure application of the law to AI systems.
- The need to ensure a level-playing field for private sector participants.

- The importance of involving stakeholders from developing countries and to consider how to avoid expanding the gap between the haves and have nots.
- That AI does not suffer and therefore does not have to be protected with rights.

SESSION 10: WRAP UP AND NEXT STEPS

Wonki Min introduced the session to focus on short summaries of each panel's discussion, the key policy challenges, possible policy solutions, and how best to facilitate international dialogue. This was followed by the stakeholder groups sharing their perspective. The session focused on key opportunities and challenges presented by AI and the respective roles of industry self-regulation, policy interventions, multi-stakeholder co-operation, and international co-operation.

Kenneth Cukier introduced five general themes from the session on "The State of AI Research": *i)* overall, machines being likely to augment rather than replace humans; *ii)* the need for algorithm safety and for "provably benevolent AI"; *iii)* the need for robots with human senses to serve humans safely; *iv)* the broader AI ecosystem and need for interoperability among AI systems; and, *v)* creating synthetic data for machine learning algorithm based on models of the world and the possible need for a "CERN for AI". From a policy perspective, he pointed out that recent advances in AI technology allow us to leverage data in fundamentally new ways compared to a few years ago, leading to the need to rethink public policies in an open-minded manner in view of the new technologies.

Andrew Wyckoff summarised session 2 on "AI Applications and Case Studies". He cited Garry Kasparov who said that AI is a tool to expand human reach, power, and knowledge and that as its power grows, so must human responsibility. He presented AI as a general purpose technology tool for applications as diverse as building smart-cities in China, powering BMW driverless cars, or tracking the sensor data of cows. He recalled that as with previous GPTs, we should expect some turbulence and disruption as well as policy challenges including; whether all firms, including SMEs, can navigate the transition; equity and whether AI will widen the gap between north and south countries, between SMEs and large businesses and between people, or provide leapfrogging opportunities. Google provides curated database, codes and training in the public domain as a way to diffuse AI. A platform like a worldwide recognised public research infrastructure CERN may share the use of curated databases and open-source programmes as well as provide trainings and this would help to alleviate the problem.

Claire Jolly summarised session 3 *"Close-up on AI in Space Applications"* that discussed emerging geospatial applications enabled by the combination of satellite data and machine learning. She reported that start-ups are providing new and lucrative ways to monitor entire sectors with satellite and other data. The Australian government is leveraging geospatial applications to monitor an entire continent and the European Commission explained that a deluge of satellite data is still to come from new constellation satellites. Policy challenges identified included: *i)* open data policies, notably for SMEs to build up innovative value-added products and services; *ii)* the need for more interoperability and standardisation of data formats from satellites; and *iii)* the strong upcoming competition to hire people with adequate skills, especially for SMEs.

Dominique Guellec summarised session 4 *"Enhancing Discovery – The role of AI in Science"* that discussed the promise of using AI in science, with automated labs, the exploitation of enormous quantities of data and using "data as the model". He reported that in practice, AI in science is used to sort and clean enormous data sets. AI complements scientists, with humans in charge of conceptual thinking such as building the research framework and setting the

context for specific experiments. Policy challenges include the need to: *i)* adapt education and train scientists to better complement AI *i.e.* to help scientists focus on conceptual thinking; and *ii)* to consider the appropriate level of government involvement in AI research to address societal grand challenges, although businesses have become dominant in AI research over the past 5 years.

Anne Carblanc summarised session 5 on the *"AI Policy Landscape"*. Non-government organisations such as the Partnership on AI and the Future Society are working to fashion the enabling environment for AI to thrive in an ethical manner. The panel showcased national initiatives to leverage AI for competitiveness (*e.g.* in China, the European Commission, Finland and France) as well as initiatives to address challenges posed by AI (*e.g.* Japan, Estonia, France or the G7). Session findings included the need: *i)* to ensure "human-centric AI", maximising benefits while addressing ethical risks, risks to privacy and transparency; *ii)* to acknowledge differences in national cultures, legal systems, country sizes and level of AI adoption; and *iii)* for multi-stakeholder collaboration to lower barriers of access to AI and data (*e.g.* with a "CERN for AI" collaboration platform and "Data Commons") and to develop flexible guidelines for AI research and applications.

Marc Keese summarised session 6 on "Employment and Skills", focusing on the relationship between humans and machines. He noted findings that half of the people in OECD member countries have a level of literacy and numeracy that AI already has. He noted that his session suggested a link between AI and job polarisation, the rise of populism and increasing inequality. He said that the challenges of work displacement by technology and globalisation and increasing job polarisation as middle class jobs disappear and demand for lower wage jobs increases were not only related to AI. He also noted the opportunities that AI brings for workers to make work more interesting by automating routine tasks, allowing more flexible work and possibly better work-life balance. He emphasised that: *i)* rather than react to specific technology developments, policy makers should evolve their policies to shape the future of work and empower individual workers through access to training and social security protection; *ii)* we need to gather better evidence through projects such as Going Digital and Future of Work; and *iii)* there is an opportunity to federate the OECD's activities and establish multi-disciplinary policy debate on AI.

Katarina de Brisis summarised Session 7 on "Privacy and Security", focusing on: preventing unwanted profiling, correcting algorithmic biases, inserting human values into automated decision making, and operationalising algorithm transparency or "explainability". She called for further discussion among stakeholders on networked platforms applying AI today and what needs to be regulated and how. She noted that AI relies on data, increasing parts of which may be assigned as personal data (*e.g.* geolocation data from sensors). She highlighted the need to monitor the impact of and ensure compliance with legislation such as GDPR, but also to design privacy measures that ensure privacy on the ground with multi-disciplinary teams embedding privacy into AI solutions and conducting privacy impact assessments to balance privacy against functionality and flexibility of the technology. She recalled the relevance of the OECD's ongoing work on principles for enhanced access to data to balance access to data to benefit society and help solve societal challenges with basic tenets of privacy.

Wonki Min summarised session 8 on *"Safety, Responsibility & Liability"* that examined the relevance and effectiveness of existing product safety and liability regimes when applied to AI and connected products. He said that there was a consensus on the need for a pragmatic, inclusive and multi-stakeholder policy debate on how AI impacts existing safety and liability concepts such as "product", "safety", "defect", and "damage", based on current and foreseeable AI technology. He noted that there is likely not a one-size-fits-all solution for autonomous

systems: for example, driverless cars seem to clearly challenge existing concepts of fault-based liability and strict liability and insurance will likely play a key role. He also stressed the need to consider the safety benefits of AI products, along with the risks and to level the standards between AI-embedded products and non-AI products.

Douglas Franz summarised session 9 on "Transparency, Oversight & Ethics" that found consensus on the urgent need for a broad based conversation with many stakeholders about the future of AI and its implementation and on the need to ensure trust, equality, transparency and accountability. He noted the need for further discussion on the most appropriate governance models for AI and their possible complementarity – from standards and self-regulation to soft-law to regulation. He highlighted the opportunities to build on the valuable existing knowledge and to distil the best principles for public policy and international cooperation. Citing for example the IEEE's Ethically Aligned Design and standards series, the Asilomar principles, and the guidelines developed by the Japanese Ministry of Internal Affairs and Communication, he stressed the need for the OECD to take action to translate insights into actionable policy principles.

Marc Rotenberg, representative of the OECD Civil Society Information Society Advisory Council (CSISAC) and President of the Electronic Privacy Information Center (EPIC), noted that the discussion on AI is not simply about those who design AI programmes and build companies, but the much broader impact of the technology on the public and on political and social institutions. He said that the current path was perilous and unsustainable due to increasing polarisation and wealth inequality partly fuelled by technology. He called on the OECD as a membership organisation of advanced democratic countries to address public challenges linked to AI and algorithm transparency to establish democratic accountability over innovation, building on conversations that began with the Global Knowledge Forum in Tokyo in 2014. Saying that AI, singularity and artificial general intelligence are accelerating processes he stressed the urgency of finding solutions. He quoted Edison "what men create with his hands, he must control its head", cautioning that consequences of inaction would be significant.

Anna Byhovskaya, Policy Advisor, Trade Union Advisory Committee to the OECD (TUAC), said that Trade unions should be part of the multi-stakeholder policy dialogue on AI, beyond just employment and skills: firm level, society-level impacts on people and people's concerns. She said that the real world outcomes of AI call for proactive policy and ex-ante regulation. She stressed that: *i)* AI should be under human command; *ii)* discussions on employment should discuss job creation and how AI will apply to the workplace and change work tasks. She said that the OECD is well placed to facilitate such discussions because it has the data and involvement of stakeholders and social partners.

She also sought stronger policy responses that do not individualise the responsibility to deal with unemployment, job changes and wage reduction especially for individuals and workers at lower skill levels, and that create job transition frameworks for workers and discussion about the financing and governance of the system and consultation from social institutions. She stressed the need for discussion on real impacts of AI, labour movement concerns about market concentration, data ownership and cost savings mechanisms in the business models as well as employer mechanisms to develop worker skills *e.g.* collective bargaining instruments, increasing productivity versus workers' autonomy, and creating rules about collection of workers' data.

Nicole Primmer, Senior Policy Director, Business at OECD (BIAC), highlighted the promises of AI but also challenges due to silos between various parts of government, governance oversight, and balancing access to data with privacy. She emphasised: *i)* the necessary partnership between business, government, technical community, civil society, and

other interested stakeholders; *ii)* discussion on skills in the age of AI; *iii)* smart policies to foster innovation and deployment of AI and encourage responsible application of AI, building on existing frameworks; *iv)* leverage AI across sectors, not forgetting SMEs; *v)* distinguish between short, medium, and long term issues related to AI. She also highlighted that AI was a core pillar of the work of the B20 under the German G20 presidency, with the B20 supporting OECD's work in this area: *i)* public dialogue on AI opportunities and challenges, *ii)* favourable ecosystems for AI, and *iii)* facilitating smart infrastructure. BIAC supports OECD work on AI because of its cross-disciplinary capacity, evidence based approach, and dialogue-based approach.

Clara Neppel, Senior Director, IEEE European Office, Internet Technical Advisory Committee (ITAC) representative, introduced the IEEE's activities on standardisation and collaboration to "Advance technology for the benefit of humanity" with its 420 000 members worldwide *(presentation)*. She said that AI profiles people based on training and on data collected to make determinations and suggestions which should be beneficial to people. Noting that for personal data to be controlled by users, agents that know our preferences could negotiate complex data sharing with other intelligent systems. She highlighted the need for inclusive and diverse AI data sets that can expand user's profiled interest or opportunities. Noting that open access may be encouraged for some data but not all data, she underscored the need for different licencing mechanisms for data, modelled on *e.g.* FRAND terms (Fair, Reasonable And Non-Discriminative) and for standardised data exchange formats.

She also stressed the need to standardise and certify algorithms, particularly for security and safety relevant systems, but also for systems that provide decisions that impact individuals' life or health. Although patents are being filed for AI systems, self-learning systems that adapt to their environment may call for rethinking AI system certification instruments and intellectual property rights mechanisms. Bigger discussion around "trust" also encompasses the issues around digital identity, digital inclusion and ethics. When we talk about ethics, it should be sought at individual level, education, as well as embedded values in the system design, and widening the measure of success beyond profit to include people and planet.

Joanna Bryson said that policy should be based on what current AI can do but emphasised the rapid pace of change of the current state of AI. She noted that humans change much more slowly and created ethics to protect ourselves she emphasised human-centric AI, *i.e.* building systems that are safe and backed up, and that people do not have worry about.

The discussion focused on next steps for the OECD on AI following the November 2016 Technology Foresight Forum on AI, the publication of the 2017 OECD "Digital Economic Outlook", and the conference. Discussants called for multi-disciplinary coordination involving the groups that focus on digital economy, science, business models, consumer protection, education and employment, with a view to developing soft-law principles in the form of a council recommendation, aiming for guidelines that can stand the test of time and are sufficiently flexible to not hinder the development of AI. The OECD was also asked to help support discussions of G7 and G20 higher level groups.

Annex. Conference Agenda

Conference Chair: **Wonki Min**, Chairman, OECD Committee on Digital Economy Policy (CDEP), Korea

Thursday 26 October 2017 - AI DEVELOPMENTS & APPLICATIONS, CC 1

9:00 – 9:20 **KEYNOTE AND WELCOME REMARKS**

Garry Kasparov, Former World Chess Champion and author of 'Deep Thinking' (by video)
Andrew Wyckoff, STI Director, OECD
Masahiko Tominaga, Vice-Minister for Policy Coordination, Ministry of Internal Affairs and Communications (MIC), Japan

9:20 – 10:50 **1. THE STATE OF AI RESEARCH**

Session moderator: **Kenneth Cukier**, Senior Editor, The Economist, United Kingdom
Francesca Rossi, Research Scientist, IBM Watson and Professor of Computer Science, University of Padova, Italy
Stuart Russell, Professor of Computer Science, University of California, Berkeley, United States
Rodolphe Gelin, Robotics Software Engineering Lead, SoftBank Robotics, Paris
Osamu Sudoh, Professor, University of Tokyo Interfaculty Initiative in Information Studies, Japan
Philipp Slusallek, Scientific Director at DFKI, Germany

10:50 – 11:20 **COFFEE BREAK AND INTERACTIVE DEMONSTRATIONS BY GOOGLE AND FACEBOOK, ATRIUM**

11:20 – 12:40 **2. AI APPLICATIONS AND CASE STUDIES**

Session moderator: **Andrew Wyckoff**, STI Director, OECD
Valerio Dilda, Partner, Paris, McKinsey & Company
Reinhard Stolle, Department of Artificial Intelligence at BMW AG, Munich
Max Yuan, founder and chairman, Xiaoi Robot Technology, Shanghai
Lynette Webb, Senior Manager, European Policy Strategy, Google, London

12:40 – 13:40 **LUNCH BREAK** (Salles Roger Okrent & George Marshall, Château de la Muette)

13:40 – 15:00 **3. CLOSE-UP ON AI IN SPACE APPLICATIONS**

Session moderator: **Claire Jolly**, Head of the OECD Space Forum
Tugdual Ceillier, Lead Data Scientist, EarthCube, Toulouse
Bryan Yates, Director of Sales - EMEA region, Orbital Insight, Mountain View, California
Thanh-Long Huynh, CEO, Quantcube Technology, Paris
Bahaa Alhaddad, Space Business Development, Starlab Space, Harwell Oxford, United Kingdom
Alexander Cooke, Counsellor, Department of Industry, Innovation and Science, Australia
Christophe Roeland, Policy Officer, Space Data for Societal Challenges and Growth, DG GROW, EC, Brussels

15:00 – 16:20 **4. ENHANCING DISCOVERY – THE ROLE OF AI IN SCIENCE**

Session moderator: **Dominique Guellec**, Head of OECD Science and Technology Policy Division
Stephen Roberts, Professor of Machine Learning in Information Engineering, University of Oxford, United Kingdom
Hiroaki Kitano, President and CEO of Sony Computer Science Laboratories, Japan
Ross King, Professor of Machine Intelligence, Manchester University School of Computer Science, United Kingdom
Jonathan McLoone, Technical Director, Wolfram Research Europe

16:20-16:50 **COFFEE BREAK AND INTERACTIVE DEMONSTRATIONS BY GOOGLE AND FACEBOOK, ATRIUM**

16:50 – 18:30 **5. AI POLICY LANDSCAPE**

Session moderator: **Anne Carblanc**, Head of OECD Division on Digital Economy Policy
David Heiner, Strategic Policy Advisor at Microsoft, representative of the Partnership on Artificial Intelligence
Nicolas Miailhe, Director for Artificial Intelligence, The Future Society @ Harvard Kennedy School of Government
ITALY: Benedetta Arese Lucini, Italy
JAPAN: Susumu Hirano, Faculty of Policy Studies / Professor, Dean, Graduate School of Policy Studies, Chuo University
FRANCE: Cédric Villani, député LREM de l'Essonne, chargé de mission IA
CHINA: Xiao Zhang, Vice Director, China Internet Network Information Center
FINLAND: Pekka Sivonen, Director, Digitalisation Strategy and Programmes, Tekes Innovation Funding Agency, Finland
ESTONIA: Marten Kaevats, National Digital Advisor, Government Office of Estonia
EUROPEAN COMMISSION: Cécile Huet, Deputy Head of Unit, Robotics and Artificial Intelligence, DG CONNECT, Brussels

18:30 **COCKTAIL RECEPTION** (Salles Roger Okrent & George Marshall, Château de la Muette)

Friday 27 Oct. 2017 - PUBLIC POLICY CONSIDERATIONS RAISED BY AI, CC12	
9:30 – 11:00	**6. EMPLOYMENT & SKILLS** Session moderator: **Mark Keese**, Head of OECD Division on Skills and Employability **Frank Levy**, Rose Professor Emeritus, MIT **Christina Colclough**, Senior Policy Advisor , UNI Global Union **James Hairston**, Head of Public Policy, Oculus VR, Facebook **Stuart Elliott**, Director of Board on Testing and Assessment, United States National Academy of Science **Young Tae Kim**, Secretary General, International Transport Forum (ITF)
11:00 – 11:30	**COFFEE BREAK AND INTERACTIVE DEMONSTRATIONS BY GOOGLE AND FACEBOOK, ATRIUM**
11:30 – 12:30	**7. PRIVACY & SECURITY** Session moderator: **Katarina de Brisis**, Deputy Director General at Ministry of Local Government and Modernisation, Norway, Chair of OECD Working Party on Security and Privacy in the Digital Economy **Peter Fleischer**, Global Privacy Counsel, Google **Taylor Owen**, Assistant Professor of Digital Media and Global Affairs, University of British Columbia **Mathias Cellarius**, Data Protection and Privacy Officer, SAP **Kenneth Cukier**, Senior Editor, The Economist, United Kingdom
12:30 – 14:00	**LUNCH BREAK** (Lunch not provided, several options available at the site and nearby)
14:00 – 15:00	**8. SAFETY, RESPONSIBILITY & LIABILITY** Session moderator: **Wonki Min**, Chairman, OECD CDEP, Korea **Rod Freeman**, international products lawyer, Partner at Cooley, United Kingdom **Hans Ingels**, Head of Unit, Single Market Policy, Mutual Recognition and Surveillance, EC, DG GROW **Pierre Chalançon**, Chair of BIAC Consumer Task Force & VP Regulatory Affairs, Vorwerk & Co KG, EU Representation **Georg Borges**, Professor, Faculty of Law, Saarland University, Germany
15:00 – 16:00	**9. TRANSPARENCY, OVERSIGHT & ETHICS** Session moderator: **Douglas Frantz**, Deputy Secretary-General, OECD **Konstantinos Karachalios**, Managing Director of the IEEE-Standards Association Joanna Bryson, Reader at University of Bath & Affiliate, Center for Information Technology Policy at Princeton University **Carolyn Nguyen**, Director of Technology Policy, Microsoft **Seán Ó hÉigeartaigh**, Executive Director of Cambridge's Centre for the Study of Existential Risk
16:00 – 16:30	**COFFEE BREAK AND INTERACTIVE DEMONSTRATIONS BY GOOGLE AND FACEBOOK, ATRIUM**
16:30 – 18:00	**10. WRAP-UP AND NEXT STEPS** Chair: **Wonki Min**, CDEP Chair BRIEF REPORTS BY SESSION MODERATORS , FOLLOWED BY STAKEHOLDER PRESENTATIONS AND DISCUSSION **Marc Rotenberg**, representative of OECD Civil Society Information Society Advisory Council (CSISAC) and President, Electronic Privacy Information Center (EPIC) **Anna Byhovskaya**, Policy Advisor, Trade Union Advisory Committee to the OECD (TUAC) **Nicole Primmer**, Senior Policy Director, Business at OECD (BIAC) **Clara Neppel**, Senior Director, IEEE European Office, Internet Technical Advisory Committee (ITAC) representative