Artificial Intelligence and Machine Learning: Policy Paper

April 2017

Foreword

Artificial intelligence (AI) is a technology that is already impacting how users interact with, and are affected by, the Internet. In the near future, its impact is likely to only continue to grow. AI has the potential to vastly change the way that humans interact, not only with the digital world, but also with each other, through their work and through other socio-economic institutions – for better or for worse.

If we are to ensure that the impact of artificial intelligence will be positive, it will be essential that all stakeholders participate in the debates surrounding AI.

In this paper, we seek to provide an introduction to AI to policymakers and other stakeholders in the wider Internet ecosystem.

The paper explains the basics of the technology behind AI, identifies the key considerations and challenges surrounding the technology, and provides several high-level principles and recommendations to follow when dealing with the technology.

If more stakeholders bring their points of view and expertise to the discussions surrounding AI, we are confident that its challenges can be addressed and the vast benefits the technology offers can be realized.

Executive Summary

Artificial Intelligence (AI) is a rapidly advancing technology, made possible by the Internet, that may soon have significant impacts on our everyday lives. AI traditionally refers to an artificial creation of human-like intelligence that can learn, reason, plan, perceive, or process natural language[1]. These traits allow AI to bring immense socio-economic opportunities, while also posing ethical and socio-economic challenges.

As AI is an Internet enabled technology, the Internet Society recognizes that understanding the opportunities and challenges associated with AI is critical to developing an Internet that people can trust.

This policy paper offers a look at key considerations regarding AI, including a set of guiding principles and recommendations to help those involved in policy making make sound decisions. Of specific focus is machine learning, a particular approach to AI and the driving force behind recent developments. Instead of programming the computer every step of the way, machine learning makes use of learning algorithms that make inferences from data to learn new tasks.

As machine learning is used more often in products and services, there are some significant considerations when it comes to users' trust in the Internet. Several issues must be considered when addressing AI, including, socio-economic impacts; issues of transparency, bias, and accountability; new uses for data, considerations of security and safety, ethical issues; and, how AI facilitates the creation of new ecosystems.

At the same time, in this complex field, there are specific challenges facing AI, which include: a lack of transparency and interpretability in decision-making; issues of data quality and potential bias; safety and security implications; considerations regarding accountability; and, its potentially disruptive impacts on social and economic structures.

In evaluating the different considerations and understanding the various challenges, the Internet Society has developed a set of principles and recommendations in reference to what we believe are the core "abilities"[2] that underpin the value the Internet provides.

While the deployment of AI in Internet based services is not new, the current trend points to AI as an increasingly important factor in the Internet's future development and use. As such, these guiding principles and recommendations are a first attempt to guide the debate going forward. They include: ethical considerations in deployment and design; Ensuring the "Interpretability" of AI systems; empowering the consumer; responsibility in the deployment of AI systems; ensuring accountability; and, creating a social and economic environment that is formed through the open participation of different stakeholders.

[1] See: https://en.wikipedia.org/wiki/Artificial_intelligence

[2] For a full list of the abilities and principles that guide our work, see http://www.internetsociety.org/who-we-are/mission/values-and-principles

Introduction

Artificial intelligence (AI) has received increased attention in recent years. Innovation, made possible through the Internet, has brought AI closer to our everyday lives. These advances, alongside interest in the technology's potential socio-economic and ethical impacts, brings AI to the forefront of many contemporary debates. Industry investments in AI are rapidly increasing[3], and governments are trying to understand what the technology could mean for their citizens[4].

The collection of "Big Data" and the expansion of the Internet of Things (IoT), has made a perfect environment for new AI applications and services to grow. Applications based on AI are already visible in healthcare diagnostics, targeted treatment, transportation, public safety, service robots, education and entertainment, but will be applied in more fields in the coming years. Together with the Internet, AI changes the way we experience the world and has the potential to be a new engine for economic growth.

Current Uses of AI:

Although artificial intelligence evokes thoughts of science fiction, artificial intelligence already has many uses today, for example:

- Email filtering: Email services use artificial intelligence to filter incoming emails. Users can train their spam filters by marking emails as "spam".
- Personalization: Online services use artificial intelligence to personalize your experience. Services, like Amazon or Netflix, "learn" from your previous purchases and the purchases of other users in order to recommend relevant content for you.
- Fraud detection: Banks use artificial intelligence to determine if there is strange activity on your account. Unexpected activity, such as foreign transactions, could be flagged by the algorithm.
- Speech recognition: Applications use artificial intelligence to optimize speech recognition functions. Examples include intelligent personal assistants, e.g. Amazon's "Alexa" or Apple's "Siri".

The Internet Society recognizes that understanding the opportunities and challenges associated with AI is critical to developing an Internet that people trust. This is particularly important as the Internet is key for the technology behind AI and is the main platform for its deployment; including significant new means of interacting with the network. This policy paper offers a look at the key things to think about when it comes to AI, including a set of guiding principles and recommendations to help make sound policy decisions. Of particular focus is machine learning, a specific approach to AI and the driving force behind recent developments.

3 See https://www.weforum.org/agenda/2016/06/investors-are-backing-more-AI-startups-than-ever-before

4 See for example: UK House of Commons Science and Technology Committee's *"Robotics and artificial intelligence"*, or White House Reports *"Preparing for the Future of Artificial Intelligence"* or *"Artificial Intelligence, Automation, and the Economy"*.

Artificial Intelligence - What it's all about

Artificial intelligence (AI) traditionally refers to an artificial creation of human-like intelligence that can learn, reason, plan, perceive, or process natural language[5].

Artificial intelligence is further defined as "narrow AI" or "general AI". Narrow AI, which we interact with today, is designed to perform specific tasks within a domain (e.g. language translation). General AI is hypothetical and not domain specific, but can learn and perform tasks anywhere. This is outside the scope of this paper. This paper focuses on advances in narrow AI, particularly on the development of new algorithms and models in a field of computer science referred to as ***machine learning***.

Machine learning – Algorithms that generate Algorithms
Algorithms are a sequence of instructions used to solve a problem. Algorithms, developed by programmers to instruct computers in new tasks, are the building blocks of the advanced digital world we see today. Computer algorithms organize enormous amounts of data into information and services, based on certain instructions and rules. It's an important concept to understand, because **in machine learning, learning algorithms – not computer programmers– create the rules.**

Instead of programming the computer every step of the way, this approach gives the computer instructions that allow it to learn from data without new step-by-step instructions by the programmer. This means computers can be used for new, complicated tasks that could not be manually programmed. Things like photo recognition applications for the visually impaired, or translating pictures into speech.[6]

The basic process of machine learning is to give *training data* to a *learning algorithm*. The learning algorithm then generates a new set of rules, based on inferences from the data. This is in essence generating a new algorithm, formally referred to as the machine learning model. By using different training data, the same learning algorithm could be used to generate different models. For example, the same type of learning algorithm could be used to teach the computer how to translate languages or predict the stock market.

Inferring new instructions from data is the core strength of machine learning. It also highlights the critical role of data: the more data available to train the algorithm, the more it learns. In fact, many recent advances in AI have not been due to radical innovations in learning algorithms, but rather by the enormous amount of data enabled by the Internet.

5 See: https://en.wikipedia.org/wiki/Artificial_intelligence

6 See the example of Aipoly, a smartphone app designed to assist people with visual impairment to navigate or identify objects by taking a photo and getting an audio description in return, https://techcrunch.com/2015/08/17/aipoly-puts-machine-vision-in-the-hands-of-the-visually-impaired/

How machines learn:

Although a machine learning model may apply a mix of different techniques, the methods for learning can typically be categorized as three general types:

- Supervised learning: The learning algorithm is given labeled data and the desired output. For example, pictures of dogs labeled "dog" will help the algorithm identify the rules to classify pictures of dogs.
- Unsupervised learning: The data given to the learning algorithm is unlabeled, and the algorithm is asked to identify patterns in the input data. For example, the recommendation system of an e-commerce website where the learning algorithm discovers similar items often bought together.
- Reinforcement learning: The algorithm interacts with a dynamic environment that provides feedback in terms of rewards and punishments. For example, self-driving cars being rewarded to stay on the road.

Why now?

Machine learning is not new. Many of the learning algorithms that spurred new interest in the field, such as neural networks[7], are based on decades old research[8]. The current growth in AI and machine learning is tied to developments in three important areas:

- Data availability: Just over 3 billion people are online with an estimated 17 billion connected devices or sensors.[9] That generates a large amount of data which, combined with decreasing costs of data storage, is easily available for use. Machine learning can use this as training data for learning algorithms, developing new rules to perform increasingly complex tasks.
- Computing power: Powerful computers and the ability to connect remote processing power through the Internet make it possible for machine-learning techniques that process enormous amounts of data[10].
- Algorithmic innovation: New machine learning techniques, specifically in layered neural networks – also known as "deep learning" – have inspired new services, but is also spurring investments and research in other parts of the field.[11]

7 Neural networks is a computational approach modeled on the human brain.

8 The history of neural networks is often described as starting with a seminal paper in 1943 by Warren McCulloch and Walter Pitts on how neurons might work, and where they modeled a simple neural network with electrical circuits.

9 See https://www.forbes.com/sites/louiscolumbus/2016/11/27/roundup-of-internet-of-things-forecasts-and-market-estimates-2016/#67bdbe2e292d

10 Google's ground breaking experiment "AlphaGo", the first AI to beat the human champion at the board game Go used approximately 280 GPU cards and 1,920 standard processors. See http://www.economist.com/news/science-and-technology/21694540-win-or-lose-best-five-battle-contest-another-milestone

11 A good example of such progress is the program Libratus, the first AI to beat several of the top human players in no-limit Texas Hold 'Em poker- a game that has been notoriously difficult for an AI to win due to incomplete information about the game state. See https://www.wired.com/2017/02/libratus/

Key Considerations

As machine learning algorithms are used in more and more products and services, there are some serious factors must be considered when addressing AI, particularly in the context of people's trust in the Internet:

- **Socio-economic impacts.** The new functions and services of AI are expected to have significant socio-economic impacts. The ability of machines to exhibit advanced cognitive skills to process natural language, to learn, to plan and to perceive, makes it possible for new tasks to be performed by intelligent systems, sometimes with more success than humans[12]. New applications of AI could open up exciting opportunities for more effective medical care, safer industries and services, and boost productivity on a massive scale.
- **Transparency, bias and accountability**. AI-made decisions can have serious impacts in people's lives. AI may discriminate against some individuals or make errors due to biased training data. How a decision is made by AI is often hard to understand, making problems of bias harder to solve and ensuring accountability much more difficult.
- **New uses for data**. Machine learning algorithms have proved efficient in analyzing and identifying patterns in large amounts of data, commonly referred to as "Big Data". Big Data is used to train learning algorithms to increase their performance. This generates an increasing demand for data, encouraging data collection and raising risks of oversharing of information at the expense of user privacy.
- **Security and safety.** Advancements in AI and its use will also create new security and safety challenges. These include unpredictable and harmful behavior of the AI agent, but also adversarial learning by malicious actors.
- **Ethics.** AI may make choices that could be deemed unethical, yet also be a logical outcome of the algorithm, emphasizing the importance to build in ethical considerations into AI systems and algorithms.
- **New ecosystems.** Like the impact of mobile Internet, AI makes new applications, services, and new means of interacting with the network possible. For example, through speech and smart agents, which may create new challenges to how open or accessible the Internet becomes.

Challenges

Many factors contribute to the challenges faced by stakeholders with the development of AI, including:

- **Decision-making: transparency and "interpretability".** With artificial intelligence performing tasks ranging from self-driving cars to managing insurance payouts, it's critical we understand decisions made by an AI agent. But transparency around algorithmic decisions is sometimes limited by things like corporate or state secrecy or technical literacy. Machine learning further complicates this since the internal decision logic of the model is not always understandable, even for the programmer[13].

12 For example, the use of machine learning for medical image analysis, https://www.technologyreview.com/s/602958/an-ai-ophthalmologist-shows-how-machine-learning-may-transform-medicine/

13 Burrell, J. *"How the machine 'thinks': Understanding opacity in machine learning algorithms".* (2016), http://journals.sagepub.com/doi/abs/10.1177/2053951715622512

While the learning algorithm may be open and transparent, the model it produces may not be. This has implications for the development of machine learning systems, but more importantly for its safe deployment and accountability. There is a need to understand why a self-driving car chooses to take specific actions not only to make sure the technology works, but also to determine liability in the case of an accident.

- **Data Quality and Bias.** In machine learning, the model's algorithm will only be as good as the data it trains on – commonly described as "garbage in, garbage out". This means biased data will result in biased decisions. For example, algorithms performing "risk assessments" are in use by some legal jurisdictions in the United States to determine an offenders risk of committing a crime in the future. If these algorithms are trained on racially biased data, they may assign greater risk to individuals of a certain race over others.[14] Reliable data is critical, but greater demand for training data encourages data collection. This, combined with AI's ability to identify new patterns or re-identify anonymized information, may pose a risk to users' fundamental rights as it makes it possible for new types of advanced profiling, possibly discriminating against particular individuals or groups.

The problem of minimizing bias is also complicated by the difficulty in understanding how a machine learning model solves a problem, particularly when combined with a vast number of inputs. As a result, it may be difficult to pinpoint the specific data causing the issue in order to adjust it. If people feel a system is biased, it undermines the confidence in the technology.

- **Safety and Security.** As the AI agent learns and interacts with its environment, there are many challenges related to its safe deployment. They can stem from unpredictable and harmful behavior, including indifference to the impact of its actions. One example is the risk of "reward hacking" where the AI agent finds a way of doing something that might make it easier to reach the goal, but does not correspond with the designer's intent, such as a cleaning robot sweeping dirt under a carpet[15].

The safety of an AI agent may also be limited by how it learns from its environment. In reinforcement learning this stems from the so-called exploration/exploit dilemma. This means an AI agent may depart from a successful strategy of solving a problem in order to explore other options that could generate a higher payoff[16]. This could have devastating consequences, such as a self-driving car exploring the payoff from driving on the wrong side of the road.

There is also a risk that autonomous systems are exploited by malicious actors trying to manipulate the algorithm. The case of "Tay", a chatbot deployed on Twitter to learn from interactions with other users, is a good example. It was manipulated through a coordinated attack by Twitter users, training it to engage in racist behavior[17]. Other

14 See, https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.bwuhXcwqn

15 Amodei, Dario, et al. *"Concrete problems in AI safety."* (2016), https://arxiv.org/abs/1606.06565

16 Ibid

17 See https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/ ; https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.0000u2hwzvo38fdew261y8vofjdej

examples of so-called "adversarial learning" include attacks that try to influence the training data of spam filters or systems for abnormal network traffic detection, so as to mislead the learning algorithm for subsequent exploitation[18].

The ability to manipulate the training data, or exploit the behavior of an AI agent also highlights issues around transparency of the machine learning model. Disclosing detailed information about the training data and the techniques involved may make an AI agent vulnerable to adversarial learning. Safety and security considerations must be taken into account in the debate around transparency of algorithmic decisions.

- **Accountability.** The strength and efficiency of learning algorithms is based on their ability to generate rules without step-by–step instructions. While the technique has proved efficient in accomplishing complex tasks such as face-recognition or interpreting natural language, it is also one of the sources of concern.

When a machine learns on its own, programmers have less control. While non-machine learning algorithms may reflect biases, the reasoning behind an algorithm's specific output can often be explained. It is not so simple with machine learning.

Not being able to explain why a specific action was taken makes accountability an issue. Had "Tay", the chatbot that engaged in racist behavior as mentioned in the prior section, broken a law (such as issuing criminal threats), would its programmers be held accountable? Or would the twitter users who engaged in adversarial training?

In most countries, programmers are not liable for the damages that flaws in their algorithms may produce. This is important, as programmers would likely be unwilling to innovate if they were. However, with the advancement of IoT technologies, such issues may become more immediate. As flaws in algorithms result in greater damages, there is a need for clarified liability on the part of the manufacturer, operator, and the programmer. With AI, the training data, rather than the algorithm itself, may be the problem. By obscuring the reasoning behind an algorithm's actions, AI further complicates the already difficult question of software liability. And as with many fields, it may well be liability that drives change.

- **Social and Economic Impact.** It is predicted that AI technologies will bring economic changes through increases in productivity. This includes machines being able to perform new tasks, such as self-driving cars, advanced robots or smart assistants to support people in their daily lives.[19] Yet how the benefits from the technology are distributed, along with the actions taken by stakeholders, will create vastly different outcomes for labor markets and society as a whole.

For consumers, automation could mean greater efficiency and cheaper products. Artificial intelligence will also create new jobs or increase demand for certain existing ones. But it also means some current jobs may be automated in one to two decades.

---

18 Huang, Ling, et al. *"Adversarial machine learning."* Proceedings of the 4th ACM workshop on Security and artificial intelligence. ACM, 2011. (2011), http://dl.acm.org/citation.cfm?id=2046692

19 White House Report 2016: *Artificial Intelligence, Automation and the Economy,* https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy

Some predict it could be as high as 47% of jobs in the United States.[20] Unskilled and low-paying jobs are more likely to be automated, but AI will also impact high-skilled jobs that rely extensively on routine cognitive tasks. Depending on the net-effect, this could lead to a higher degree of structural unemployment.

Automation may also impact the division of labor on a global scale. Over the past several decades, production and services in some economic sectors has shifted from developed economies to the emerging economies, largely as a result of comparatively lower labor or material costs. These shifts have helped propel some of the world's fastest emerging economies and supports a growing global middle class. But, with the emergence of AI technologies, these incentives could lessen. Some companies, instead of offshoring, may choose to automate some of their operations locally.

The positive and negative impacts of AI and automation on the labor market and the geographical division of labor will not be without their own challenges. For instance, if AI becomes a concentrated industry among a small number of players or within a certain geography, it could lead to greater inequality within and between societies. Inequality may also lead to technological distrust, particularly of AI technologies and of the Internet, which may be blamed for this shift.

- Governance. The institutions, processes and organizations involved in the governance of AI are still in the early stages. To a great extent, the ecosystem overlaps with subjects related to Internet governance and policy. Privacy and data laws are one example.

Existing efforts from public stakeholders include the UN Expert Group on Lethal Autonomous Weapons Systems (LAWS), as well as regulations like the EU's recent General Data Protection Regulation (GDPR) and the "right to explanation" of algorithmic decisions.[21] How such processes develop, and how similar regulations are adopted or interpreted, will have a significant impact on the technology's continued development. Ensuring a coherent approach in the regulatory space is important, to ensure the benefits of Internet-enabled technologies, like AI, are felt in all communities.

A central focus of the current governance efforts relates to the ethical dimensions of artificial intelligence and its implementation. For example, the Institute of Electrical and Electronics Engineers (IEEE) has released a new report on *Ethically Aligned Design* in artificial intelligence[22], part of a broader initiative to ensure ethical considerations are incorporated in the systems design. Similarly, OpenAI, a non-profit research company in California has received more than 1 billion USD in commitments to promote research and activities aimed at supporting the safe development of AI. Other initiatives from the private sector include the "Partnership on AI", established by Amazon, Google, Facebook, IBM, Apple and Microsoft "to advance public understanding of artificial intelligence technologies (AI) and formulate best practices on the challenges and opportunities within the field".

20 Frey, C. B. and Osborne, M. A. *"The Future of Employment: How Susceptible are Jobs to Computerization?"* (2013), http://www.oxfordmartin.ox.ac.uk/publications/view/1314

21 Goodman, B. and Flaxman, S. *"European Union regulations on algorithmic decision-making and a" right to explanation"* (2016), https://arxiv.org/abs/1606.08813

22 See http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Despite the complexity of the field, all stakeholders, including governments, industry and users, should have a role to play to determine the best governance approaches to AI. From market-based approaches to regulation, all stakeholders should engage in the coming years to manage the technology's economic and social impact. Furthermore, the social impact of AI cannot be fully mitigated by governing the technology, but will require efforts to govern the impact of the technology.

Guiding Principles and Recommendations

The Internet Society has developed the following principles and recommendations in reference to what we believe are the core "abilities"[23] that underpin the value the Internet provides. While the deployment of AI in Internet based services is not new, the current trend points to AI as an increasingly important factor in the Internet's future development and use. As such, these guiding principles and recommendations are a first attempt to guide the debate going forward. Furthermore, while this paper is focused on the specific challenges surrounding AI, the strong interdependence between its development and the expansion of the Internet of Things (IoT) demands a closer look at interoperability and security of IoT devices[24].

Ethical Considerations in Deployment and Design

Principle: AI system designers and builders need to apply a user-centric approach to the technology. They need to consider their collective responsibility[25] in building AI systems that will not pose security risks to the Internet and Internet users.

Recommendations:

- Adopt ethical standards: Adherence to the principles and standards of ethical considerations in the design of artificial intelligence[26], should guide researchers and industry going forward.
- Promote ethical considerations in innovation policies: Innovation policies should require adherence to ethical standards as a pre-requisite for things like funding.

Ensure "Interpretability" of AI systems

Principle: Decisions made by an AI agent should be possible to understand, especially if those decisions have implications for public safety, or result in discriminatory practices.

Recommendations:

- Ensure Human Interpretability of Algorithmic Decisions: AI systems must be designed with the minimum requirement that the designer can account for an AI agent's behaviors. Some systems with potentially severe implications for public safety should also have the functionality to provide information in the event of an accident.

23 For a full list of the abilities and principles that guide our work, see http://www.internetsociety.org/who-we-are/mission/values-and-principles

24 For more information about IoT and ISOC's guiding principles in this area, please see our previous publications:
- https://www.internetsociety.org/policybriefs/iot
- https://www.internetsociety.org/doc/iot-overview

25 ISOC, "*Collaborative Security: An approach to tackling Internet Security issues*" (2015), http://www.internetsociety.org/collaborativesecurity

26 For an example, see the principles and standards under development by the IEEE, http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

- **Empower Users:** Providers of services that utilize AI need to incorporate the ability for the user to request and receive basic explanations as to why a decision was made.

Public Empowerment
Principle: The public's ability to understand AI-enabled services, and how they work, is key to ensuring trust in the technology.

Recommendations:

- **"Algorithmic Literacy" must be a basic skill:** Whether it is the curating of information in social media platforms or self-driving cars, users need to be aware and have a basic understanding of the role of algorithms and autonomous decision-making. Such skills will also be important in shaping societal norms around the use of the technology. For example, identifying decisions that may not be suitable to delegate to an AI.
- **Provide the public with information:** While full transparency around a service's machine learning techniques and training data is generally not advisable due to the security risk, the public should be provided with enough information to make it possible for people to question its outcomes.

Responsible Deployment
Principle: The capacity of an AI agent to act autonomously, and to adapt its behavior over time without human direction, calls for significant safety checks before deployment, and ongoing monitoring.

Recommendations:

- **Humans must be in control:** Any autonomous system must allow for a human to interrupt an activity or shutdown the system (an "off-switch"). There may also be a need to incorporate human checks on new decision-making strategies in AI system design, especially where the risk to human life and safety is great.
- **Make safety a priority:** Any deployment of an autonomous system should be extensively tested beforehand to ensure the AI agent's safe interaction with its environment (digital or physical) and that it functions as intended. Autonomous systems should be monitored while in operation, and updated or corrected as needed.
- **Privacy is key:** AI systems must be data responsible. They should use only what they need and delete it when it is no longer needed ("data minimization"). They should encrypt data in transit and at rest, and restrict access to authorized persons ("access control"). AI systems should only collect, use, share and store data in accordance with privacy and personal data laws and best practices.
- **Think before you act:** Careful thought should be given to the instructions and data provided to AI systems. AI systems should not be trained with data that is biased, inaccurate, incomplete or misleading.
- **If they are connected, they must be secured:** AI systems that are connected to the Internet should be secured not only for their protection, but also to protect the Internet from malfunctioning or malware-infected AI systems that could become the next-generation of botnets. High standards of device, system and network security should be applied.
- **Responsible disclosure:** Security researchers acting in good faith should be able to responsibly test the security of AI systems without fear of prosecution or

other legal action. At the same time, researchers and others who discover security vulnerabilities or other design flaws should responsibly disclose their findings to those who are in the best position to fix the problem.

Ensuring Accountability
Principle: Legal accountability has to be ensured when human agency is replaced by decisions of AI agents.

Recommendations:

- Ensure legal certainty: Governments should ensure legal certainty on how existing laws and policies apply to algorithmic decision-making and the use of autonomous systems to ensure a predictable legal environment. This includes working with experts from all disciplines to identify potential gaps and run legal scenarios. Similarly, those designing and using AI should be in compliance with existing legal frameworks.
- Put users first: Policymakers need to ensure that any laws applicable to AI systems and their use put users' interests at the center. This must include the ability for users to challenge autonomous decisions that adversely affect their interests.
- Assign liability up-front: Governments working with all stakeholders need to make some difficult decisions now about who will be liable in the event that something goes wrong with an AI system, and how any harm suffered will be remedied.

Social and Economic Impacts
Principle: Stakeholders should shape an environment where AI provides socio-economic opportunities for all.

Recommendations:

- All stakeholders should engage in an ongoing dialogue to determine the strategies needed to seize upon artificial intelligence's vast socio-economic opportunities for all, while mitigating its potential negative impacts. A dialogue could address related issues such as educational reform, universal income, and a review of social services.

Open Governance
Principle: The ability of various stakeholders, whether civil society, government, private sector or academia and the technical community, to inform and participate in the governance of AI is crucial for its safe deployment.

Recommendations:

- Promote Multistakeholder Governance: Organizations, institutions and processes related to the governance of AI need to adopt an open, transparent and inclusive approach. It should be based on four key attributes: *Inclusiveness and transparency; Collective responsibility*; *Effective decision making and implementation* and *Collaboration through distributed and interoperable governance*[27]

27 For more information about the key attributes of multistakeholder governance, please see ISOC's *"Internet Governance - Why the Multistakeholder Approach Works"*, https://www.internetsociety.org/doc/internet-governance-why-multistakeholder-approach-works

Acknowledgments

The Internet Society acknowledges the contributions of staff members, external reviewers, and Internet Society community members in developing this paper. Special acknowledgements are due to the Internet Society's Carl Gahnberg and Ryan Polk who conducted the primary research and preparation for the paper, and Steve Olshansky who helped develop the document's strategic direction and provided valuable input throughout the writing process.

The paper benefitted from the reviews, comments and support of a set of Internet Society staff: Constance Bommelaer, Olaf Kolkman, Konstantinos Komaitis, Ted Mooney, Andrei Robachevsky, Christine Runnegar, Nicolas Seidler, Sally Wentworth and Robin Wilton. Thanks to the Internet Society Communications team for shaping the visual aspect of this paper and promoting its release: Allesandra Desantillana, Beth Gombala, Lia Kiessling, James Wood and Dan York.

Special thanks to Walter Pienciak from the Institute of Electrical and Electronics Engineers (IEEE) for his significant contributions in his early review of the paper.

Finally, the document was immensely improved by the input of a variety of Internet Society community members. Their wide areas of expertise and fresh perspectives served to greatly strengthen the final paper.