ITI's Global AI Policy Recommendations

March 2021

ITI
Promoting Innovation Worldwide

Executive Summary

Artificial Intelligence (AI) is a foundational technology that offers enormous and diverse societal benefits, including sustainability, public health and safety, cybersecurity, agriculture and economic growth. Its development and use continue to rapidly evolve. Many countries are working to harness the benefits of AI, while considering various approaches to address societal challenges that may emerge. To innovate and prosper responsibly and securely, governments need to make strategic decisions regarding AI research and development, regulation, and standards.

In an effort to help guide global governments, *ITI's Global AI Policy Recommendations* build upon its predecessor *AI Policy Principles*[1], which outline specific areas where industry, governments, communities, and other stakeholders can collaborate. *ITI's Global AI Policy Recommendations* provide globally applicable proposals on policy measures in five key areas for AI policymaking: **innovation and investment; facilitating public understanding and public trust; ensuring security and privacy; approaches to regulation;** and **global engagement.**

Innovation & Investment in AI. A robust strategy that supports multiple components of innovation and investment is necessary to harness the technology in a way that benefits many groups across society. Governments should consider actions that **guarantee a skilled and diverse workforce, invest in R&D, utilize publicly available data, and prioritize** and **enhance transparency of AI-enabled ICT procurements.**

Facilitating Public Understanding of and Trust in AI. Governments globally can play a vital role in facilitating dialogues about AI between companies that use and develop the technology and the communities they impact. Governments should **encourage a responsible design approach and promote the development of meaningfully explainable AI systems.**

Ensuring the Security & Privacy of AI Systems. Cybersecurity and privacy are foundational to trustworthy AI systems, and it is important that policymakers consider the cybersecurity implications of AI systems and ensure that users trust that their personal and sensitive data is protected and handled appropriately. **Policies should support the use of AI for cybersecurity purposes, incorporate AI systems into threat modeling and security risk management, encourage the use of global security standards, invest in security innovation to counter adversarial AI,** and **develop and support guidance that protects privacy and promotes the ethical use of data.**

Approaches to AI Regulation. Governments considering regulating AI should do so in a way that focuses on responding effectively to specific harms while allowing for advancements in technology and innovation. That includes efforts **to align around common parameters and consider the scope of AI, taking a risk-based, context-specific approach to governing AI, and evaluating existing laws and regulations in order to determine whether there are gaps requiring incremental new rules for AI.**

Global Engagement on AI. As countries around the world seek to deploy and utilize AI, collaboration and conversation is key to ensuring approaches are interoperable and aligned to the extent possible. Governments should **recognize the need for global cooperation and coordination, maintain interoperability across borders by engaging in ongoing dialogue, ensure the protected free flow of data across borders, and support global, voluntary, industry-led standards.**

Contents

Innovation & Investment in AI

Innovation and investment in AI will be key to facilitating the development and uptake of AI applications. A robust strategy that supports multiple components of innovation and investment will be necessary to harness the technology in a way that benefits many groups across society. Governments should consider actions that guarantee a skilled workforce, utilize publicly available data, and support innovation.

In order to facilitate such a climate, governments should:

1 Support policies that will help to develop a skilled and diverse AI workforce.

These policies could include modernizing candidate recruitment, hiring, and training, and should establish and advance industry-informed skilling and re-skilling programs to prepare individuals for the future of work, including an AI-enabled future. To support these initiatives, we encourage governments and businesses to continue to focus on policies designed to advance and incentivize professional and technical apprenticeships, education and training programs in STEM fields, and access to external and online reskilling programs. That said, AI is not just a function of STEM or advanced technical training; the best way to ensure access to an AI workforce is to invest broadly across all relevant disciplines and teach flexible skills and problem solving from early childhood education. At the university level, AI and/or data science programs should incorporate the social sciences, humanities, and history to integrate humanistic approaches into the curriculum beyond a single, separated "AI ethics" unit.

2 Invest in R&D for AI including basic science.

We encourage robust government support for research and development (R&D) to foster innovation in AI through R&D incentives, and increased government funding of both foundational basic science research and AI-specific research programs. As the primary source of funding for long-term, high-risk research initiatives, we support governments' investment in research fields specific or highly relevant to AI, including cyber-defense, data analytics, detection of fraudulent transactions or messages, adversarial machine learning/AI (how to secure ML/AI), privacy preserving machine learning (PPML), robotics, human augmentation, natural language processing, interfaces, and visualizations.

3 Increase access to government sources of publicly available data, as appropriate, in machine-readable formats and across borders to enable access to a foundational building block of AI.

Data is fundamental to innovation in AI. It is important that data is high-quality, credible, timely and available in machine-readable formats. By leveraging large and diverse datasets and increased computing power and ingenuity, AI developers and other stakeholders will be able to innovate and find solutions to meet the needs of individuals and society in unprecedented ways.[2] More available data means more data with which to train algorithms, resulting in higher quality AI offerings. Governments can also promote existing international standards regarding data, and data quality, and promote the development of new standards for data quality. In addition to making data available, governments may be able to curate widely available data as labeled, diverse, representative, quality data for the purposes of training corresponding AI.

4 Enhance transparency of AI-enabled ICT procurements across government agencies and ministries.

Transparency in AI-enabled ICT procurements can provide one metric to identify AI champions within government. Laggers may not know how best to incorporate AI and would benefit from the sharing of best practices and use-cases. For example, adoption leaders could host inter-agency fora to discuss challenges and successes or put together senior level briefings to provide other agencies with a model roadmap for AI adoption. Additionally, the aggregated information provides insights on how to allocate or redistribute available resources more efficiently to promote the adoption of AI technologies.

5 Prioritize procurement of AI-based technologies and applications as part of IT modernization efforts.

The deployment of AI tools will help governments at all levels leverage data generated by the public sector as a strategic asset and make more informed decisions about the actions that would best serve constituents and ensure mission success. Governments should upgrade legacy systems and make ample investments in AI and similar technologies like robotic processing automation (RPA). Modern cyberthreats are increasingly automated, and machine learning-based prevention technologies are developing to increasingly leverage AI; investing in these sophisticated cybersecurity technologies should be a priority as well.

Facilitating Public Understanding of and Trust in AI

One of the most important ways governments can foster the adoption of AI technology is to facilitate public understanding and trust. Governments globally can play a vital role in facilitating dialogues about AI between companies that use and develop AI and the communities they impact, with the intent of better aligning them. As these stakeholders become more closely aligned, trust will grow, and AI adoption will scale. To grow public understanding of and trust in AI, governments should:

1 Partner with or fund university programs whereby data science and other students in aligned disciplines conduct real world projects with communities in key areas of social need..

This can significantly improve students' skills while also providing a tangible benefit to social groups in need. It also serves a training function for the communities involved, who learn what problems AI can and cannot solve, and how to make the technology work for them in a beneficial way. Some of those community members will also develop an interest in AI and might go on to work in the field. Sometimes recent STEM graduates may believe that data science is merely technical, whereas a really a portion of the job is in fact understanding the problem domain and the stakeholders involved in order to translate their needs into data formats. A program along these lines would solve that problem while also building public trust.

2 Consider how to best promote the development of meaningfully explainable[3] AI systems.

Explainability is important for developing accountable and trustworthy AI because it enables the next step, agency—enabling entities to make decisions to avoid negative outcomes. Together, explainability and agency can foster accountability and increase public trust. Several recommendations may assist policymakers in this effort:

- Invest in research for and promote tools to achieve appropriate levels of explainability and pursue avenues to work with local research institutions, industry, and international partners in developing a common lexicon for trustworthy AI.
- Take a risk-based approach to explainability, considering where explainability makes sense in the AI space and where it might not. Explainability, while helpful in certain cases, does not make sense in every instance and could serve to hamstring innovation. Some low-risk applications of AI, for example, may not necessitate the same type of explanations as higher-risk applications; and in more benign applications that carry non-significant impacts on individuals, explanations may not be necessary at all.
- Explainability must also be balanced in consideration of other factors, including the rights of individuals to receive an explanation, the interests of businesses to maintain trade secrets, and the potential value of exposed data to potential adversaries. In considering AI explanations, value to the consumer is key – one of the benefits of AI explanations is to help individuals understand how the use of AI will benefit them. It is also important to strike a balance so that consumers do not experience "decision fatigue" and can understand the use of the AI technology without being bogged down in technical details. Conversely, keeping some information related to AI systems obscured is important to protect businesses' intellectual property interests, as well as the security of AI systems more broadly.
- Policymakers should avoid governance that creates an environment where outliers are viewed as a flaw in an overall AI system. If an outlier is indeed an outlier, then the algorithm will learn and dismiss it in later iterations so no "explanation" is necessary. As such, when and how an "explanation" may be required is highly contingent on the stage of an AI system's developmental lifecycle, the context in which a later-stage model is deployed, the purposes for which it is deployed, and numerous other factors. Any guidelines related to transparency or explainability should capture a statistically meaningful number of results to ensure uncertain results are actual concerns and not just isolated anomalies.

3 Rely on conformity assessment only in conjunction with other tools and after a comprehensive evaluation of whether a regulatory approach is warranted.

Governments are understandably considering whether conformity assessment systems can play a role in helping to generate confidence in AI systems. Conformity assessment systems include a suite of tools that provide for a range of approaches that can be calibrated to the level of risk associated with an AI system. However, whether the use of conformity assessment is appropriate in the context of AI systems should be carefully considered, and at a minimum requires identifying whether standards exist that can be used for conducting conformity assessment related activities such as certification of products or systems. Governments should be mindful in their use of conformance approaches to ensure that they do not prescribe or mandate the use of approaches that can generate a misplaced sense of security or lead to increased costs for customers. In rapidly evolving systems such as AI systems, the use of conformity assessment must be tailored in a manner that factors in the constantly evolving nature of AI systems. A static approach where an AI system or product is tested or certified at the outset may provide little assurance about the product's operation after it has been in use. In any case, any conformity assessment scheme should be developed within already existing sectorial legislation frameworks (e.g., medical devices, motor vehicles).

4 Encourage an ethical design approach.

In designing AI systems, governments should encourage an approach that promotes fairness and non-discrimination. One set of perspectives that may be worth considering are the Guidelines developed by the European High-Level Expert Group (HLEG Guidelines), which set forth seven foundational principles that characterize a trustworthy AI system. These principles include human agency and oversight, transparency, robustness and safety, privacy and data governance, diversity, non-discrimination and fairness, societal and environmental well-being and accountability.[4] However, since not all AI applications raise ethical questions as considered in the HLEG Guidelines, considerations of this type should be context-specific and limited to high-risk applications. We expand on some key elements of the HLEG Guidelines below:

- Not all AI applications require the same level of human agency, and while for some applications it is very important (e.g., uses in aviation), it is not needed for others (e.g., baggage handling systems). Thus, when determining the degree of human involvement and oversight needed, individual use cases should be taken into account.

- Context and risk-level of different AI applications vary in terms of their impact on fundamental rights. For instance, the use of AI for automated baggage handling in airports poses no risk to fundamental human rights,[5] as opposed to applications in the field of HR.

- We recommend discerning between understandability and interpretability. Understandability enables a non-technical person (e.g., business executive or customer) to gain insight into how an algorithm works and why it made a given decision. Interpretability allows a technical expert, such as an AI/machine learning expert, to understand why an algorithm made a given decision. Both understandability and interpretability are key components of an ethical design of AI. The distinction is necessary because the technical details of an AI system are not necessarily meaningful or beneficial for the end-user.

- Where possible, techniques such as anonymization, pseudonymization, de-identification and other privacy enhancing techniques (PETs) and Privacy Preserving Machine Learning (PPML) are crucial to ensure data can be used to train algorithms and perform AI tasks without breaching privacy. Users of AI can leverage "federated learning" which means they can aggregate data in ways so that the individual data points are completely private, but AI can be performed on the aggregate with minimal loss of accuracy.

Ensuring the Security & Privacy of AI Systems

Cybersecurity and privacy are foundational to trustworthy AI systems, and there are multiple ways in which AI, cybersecurity, and privacy interact. First, AI is becoming increasingly essential to cybersecurity deterrence capabilities. We discuss this further in Annex A, where we outline a cybersecurity use case for AI. Second, it is important that policymakers consider how to ensure the cybersecurity of AI systems. Third, users must trust that their personal and sensitive data used by AI systems is protected and handled appropriately.

Government policymakers should:

1 Ensure that policies support the use of AI for cybersecurity purposes.

Cybersecurity tools and technologies should incorporate AI to keep pace with the evolving threat landscape, including attackers who are constantly improving their sophisticated and highly automated methods to penetrate organizations and evade detection. Defensive cybersecurity technology can use machine learning and AI to more effectively address today's automated, complex, and constantly evolving cyberattacks. When combined with cloud, AI can help to scale cyber efforts through smart automation and continuous learning that drives self-healing systems. To support and enable the use of AI for cybersecurity purposes, policymakers must carefully shape (or reaffirm)[6] any policies related to privacy to affirmatively allow the use of personal information such as IP addresses to identify malicious activity.

2 Encourage public and private sector stakeholders to incorporate AI systems into threat modeling and security risk management.

This should include encouraging organizations to ensure that AI applications and related systems are in scope for organizational security program monitoring and testing and that the risk management implications of AI systems as a potential attack surface are considered.

3 Encourage the use of strong, globally accepted and deployed cryptography and other security standards that enable trust and interoperability in AI systems.

The tech sector incorporates strong security features into our products and services to advance trust, including AI systems. Policymakers should promote policies that support using published algorithms as the default cryptography approach as they have the greatest trust among global stakeholders, and limit access to encryption keys.

4 Invest in security innovation to counter adversarial AII.

It is important that businesses and governments also invest in cybersecurity directed at countering adversarial AI. For example, malicious actors can use adversarial AI to cause machine learning models to misinterpret inputs into the system and behave in a way that is favorable to the attacker. To produce the unexpected behavior, attackers create "adversarial examples" that often resemble normal inputs, but instead are meticulously optimized to break the model's performance. Adversarial AI represents an incremental threat compared to traditional cyber-attacks, so it important that governments ensure their policy instruments do not inadvertently stifle industry's efforts to counter adversarial AI.

5 Develop and support frameworks and guidelines that protect privacy and promote the appropriate/ethical use of data that may be used in data sets underpinning AI.

To protect personal information and support fundamental human rights, data in data sets used by AI systems may be required to be anonymized, aggregated, or otherwise de-identified such that the datasets exclude any personal information and cannot be re-identified. Doing so ensures the beneficial use of the data in training intelligent systems while protecting individual privacy and security consistent with protecting fundamental human rights.

Approaches to Regulation

We recommend that when governments consider regulating AI, they do so in a way that focuses on responding effectively to specific harms while allowing for advancements in technology and innovation. In doing so, governments should fully evaluate regulatory and non-regulatory approaches and only proceed to regulatory approaches when absolutely necessary. Regulation should be design-neutral and risk-based. In taking such an approach, governments can help ensure that their policy levers address actual demonstrated needs and are narrowly tailored, and do not inadvertently capture unrelated AI uses. In developing regulatory approaches to AI, we offer the following recommendations:

1 Align around common parameters and consider the scope of AI.

While there is not currently a universally accepted definition of AI, policymakers should strive to coalesce around parameters of what constitutes an AI system to advance globally consistent AI policy approaches. Appendix A compiles several key AI terms around which international consensus is emerging, including AI systems, machine learning, etc. Looking beyond the various relevant definitions, the scope of AI is incredibly broad and could theoretically capture many different types of systems and processes – so carefully articulating the scope of AI implicated by a regulation is essential to establishing an informed baseline for AI policymaking.

An essential factor is to properly identify the component parts of AI systems beyond algorithms (such as datasets and computing power), as well as to define related key terms such as machine learning. Some algorithms have been applied for decades but do not constitute "artificial intelligence" or "machine learning" systems. In crafting any sort of incremental AI regulation, policymakers must be clear on what aspect of AI they are referring to and in what context. There is a difference between the latest wave of AI systems that learn from data and experience, and traditional software and control systems that operate according to predictable rules, which have long been embedded in a wide variety of high-risk systems, from flight control to pacemakers. In crafting any sort of incremental AI regulation, policymakers must be clear on what aspect of AI they are referring to and in what context.

2 Take a risk-based, context-specific approach to governing AI.

Risks need to be identified and mitigated in the context of the specific AI use. This will help policymakers determine use cases or applications that are of particular concern, avoiding overly prescriptive approaches that may serve to stifle innovation. Beyond that, and as we reference in our Facilitating Public Trust in AI section above, context is key. Not all AI applications negatively impact humans and thus inflict no harm that would warrant regulation.

We recommend that policymakers, in close consultation with industry and other stakeholders, consider how to characterize "high-risk" applications of AI, including by identifying the appropriate roles for AI developers and users in making risk determinations. In our view, an AI decision is high-risk when a negative outcome could have a significant impact on people—especially as it pertains to health, safety, freedom, discrimination, or human rights. In thinking about high-risk applications, focusing on "sectors" may lead to overly broad categorizations – it is important to use a sufficiently targeted and well-outlined classification to ensure this criterion does not become irrelevant. We encourage developing a categorization that takes into account sector, use case, complexity of the AI system, probability of worst-case occurrence, irreversibility and scope of harm in worst-case scenarios e.g., individual v. larger groups of people, and other criteria.

A risk-based, context-specific approach will be the most effective means of addressing concerns that may be associated with AI, while simultaneously allowing for innovation and agility in development of AI applications. The AI development process is fast-evolving, highly varied between organizations, and geographically and technologically diffuse. AI models themselves have the potential to be complex and highly commercially sensitive. These factors combine to suggest that an extensive, prescriptive, 'one size fits all' approach to AI governance will face similar, if not greater, challenges than in other areas of technology policy. This challenge is also manifest in the very diverse application of AI and machine learning (ML) solutions which vary in sensitivity, risk and benefit.

3 Evaluate existing laws and regulations overlapping with or adjacent to various aspects of AI in order to determine whether there are gaps requiring incremental new rules for AI.

Because AI is a horizontal technology, one should evaluate its use and impact in specific applications and evaluate whether existing rules, governing areas such as data protection/privacy or product liability already address possible emerging concerns - rather than developing technology-specific laws. AI specific laws would run counter to the principle of technological neutrality and such broad regulation would likely become obsolete and possibly disproportionate to addressing identified risks related to certain applications as technology and use cases evolve. Instead, governments should work with industry and other AI stakeholders to focus new rules on the use of technology in order to address the potential issues arising in specific uses and applications.

When conducting an evaluation, governments should first identify what laws impact the use cases they are concerned with and then proceed to evaluate whether new rules are needed. They should also avoid potential conflicts of law. Policymakers should seek to ensure that any existing or forthcoming regulatory requirements do not create unnecessary technical barriers to trade and rely on global, industry-driven, voluntary, consensus-based standards. Prior to drafting legislation or regulations, it is important that policymakers consider how relevant, established and/or developing international standards can inform the development of effective laws and regulations.

Perhaps even more important than identifying and addressing gaps is clarifying how existing laws apply to AI and how AI can be used in compliance with existing laws. Many customers, particular in highly regulated industries, are hesitant to use AI services, because there is little certainty as to how such services can/should be used in compliance with existing law. Granular, technical guidance or instruction in this regard would be very helpful for giving these types of customers confidence to use AI services

Global Engagement

As countries around the world seek deploy and utilize AI, collaboration and conversation is key to ensuring approaches are interoperable and aligned to the extent possible. The potential for regulatory divergence is great, especially if countries undertake the development and deployment of AI in a vacuum. Thus, we recommend governments:

1 Recognize the need for global cooperation and coordination.

Nations around the world seek to establish trust in AI applications. The OECD made important progress in establishing AI principles of trustworthiness and in ensuring that AI remains human-centered consistent with fundamental (democratic) values. These principles provide a constructive policy blueprint, and we encourage policymakers globally to look to them when considering how to approach AI to maintain alignment.[7] We also encourage nations to join the Global Partnership on AI (GPAI), which is an important vehicle for international collaboration. The Group is currently exploring issues related to responsible AI, data governance, the future of work, and innovation and commercialization.[8]

2 Seek to maintain interoperability across borders by engaging in ongoing dialogue.

Governments should seek to maintain global interoperability and alignment of various AI frameworks around the OECD AI principles referenced above to the greatest extent possible. In this era of global digital commerce, political and regulatory divergence poses real risks to the socio-economic benefits and opportunities of data-driven technologies such as AI, where fair, accurate, fit-for-purpose models depend on access to large, diverse data sets that can flow across borders. We believe that international dialogue will spur wider dissemination of best practices, information, and guidance, increasing the likelihood that policy and regulatory approaches are interoperable.

3 Ensure the protected free flow of data across borders.

To fully realize the benefits of AI, governments need to ensure that data and meta-data can flow freely and protected across borders. Data is and will continue to be foundational to AI. As such, we encourage governments to strengthen their commitment to facilitating the free and protected flow of data across borders and refrain from imposing localization measures.

4 Support global, voluntary, industry-led standards.

AI standards are essential to increase interoperability, harmonization, and trust in AI systems. They can inform AI regulation, practical implementation, governance and technical requirements. Governments should work to support global, voluntary, industry-led standards, and safeguard the work and processes of international standards development bodies. Broad contributions to and adoption of international standards reduces market access barriers. Standards work for the net benefit of the international community and should be developed and applied without prejudice to cultural norms and without imposing the culture of any one nation. Standards work should also be technology neutral (avoiding preferential treatment for any specific technical approach).

For example, ISO/IEC JTC 1/SC 42 on Artificial Intelligence has started working on Artificial Intelligence Management System (AIMS) standard that will cover processes with development or use of AI, such as bias, fairness, inclusiveness, safety, security, privacy, accountability, explicability, and transparency. This management system standard will help in innovation and technology development through structured governance and appropriate risk management. SC 42 currently also has other standards under development, focused variously on terminology, reference architecture, governance of AI, and trustworthiness.

Appendix A

Use Cases

AI is positively transforming industries and the digital economy in myriad ways. From healthcare to environment to financial services, AI is being harnessed by businesses to positively impact society. Some examples of these use cases follow.

Climate/Environment:

AI can contribute to solutions to help address environmental and climate issues. Using computer vision in an underwater environment, a situation in which human interference would otherwise interfere with quality results, AI can be used to help gauge reef health by analyzing fish populations and marine life in real-time. This is an example of how combining human capabilities with artificial intelligence can stretch the boundaries of what is possible and provide us with data that can enable us help inform environmental and economic policy, as well as public health.

Cybersecurity:

AI and machine learning can be leveraged to improve cybersecurity. Indeed, defensive cybersecurity technology must embrace machine learning and AI as part of the ongoing battle between attackers and defenders. The threat landscape constantly evolves, with cyberattacks that are complex, automated and constantly changing. Attackers continually improve their sophisticated and highly automated methods, moving throughout networks to evade detection. The cybersecurity industry is innovating in response: making breakthroughs in machine learning and AI to detect and block the most sophisticated malware, network intrusions, phishing attempts, and many more threats.

Other examples include using AI to identify unknown IoT devices as well as suspicious device behavior, to uncover suspicious DNS activity, and to stop incoming threats.

Healthcare:

AI applications are starting to make an impact on our healthcare system today and are poised to revolutionize everything from the back office to the doctor's office, and from the emergency room to the living room. AI technologies are being used to develop insights on large patient populations in order to help predict, pre-empt and prevent people from getting sicker; to gain a better understanding of the human genome to more precisely deliver patient care; to make medical image analysis faster and more accurate for personalized treatment; and to speed-up the development of drugs and therapies and to detect and correct waste, fraud and abuse in healthcare spending.

Financial services:

AI can be used in many ways in the financial services sector. For example, AI can be used to distill complex tax laws into a personalized decision system enabling individuals to quickly and accurately file taxes. AI can also be used to improve the risk accuracy of underwriting models. By relying on alternative data beyond credit scores, AI can expand access to capital, allowing individuals and small businesses to take out loans that they otherwise might not qualify for. Finally, AI can be used for financial forecasting, using external and personalized factors to predict cash flow or plan for certain financial scenarios. This can help businesses make more informed decisions about hiring and other operations.[9]

Appendix B

Glossary of Key Terms

Below, we compile a series of definitions around which consensus is emerging. While we do not prescribe what definition a government should adopt, we encourage an alignment around key parameters.

Artificial Intelligence or AI System

- ITI definition
  - » **AI system:** An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.[10]
- OECD Definition
  - » **Artificial Intelligence:** A suite of technologies capable of learning, reasoning, adapting, and performing tasks in ways inspired by the human mind.[11]
- U.S. Consumer Product Safety Commission
  - » **Artificial Intelligence:** Any method for programming computers or products to enable them to carry out tasks or behaviors that would require intelligence if performed by humans.[12]
- European Commission
  - » **Artificial Intelligence:** Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).[13]
- European High Level Expert Group on AI
  - » **Artificial Intelligence:** Software (and possibly also hardware) systems designed by humans 3 that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).[14]

Adversarial AI

- World Economic Forum
  - » **Adversarial AI:** the malicious development and use of advanced digital technology and systems that have intellectual processes typically associated with human behavior. These include the ability to learn from past experiences, and to reason or discover meaning from complex data.[15]

Machine-learning

- U.S. Consumer Product Safety Commission
  - » **Machine-learning:** An iterative process of applying models or algorithms to data sets to learn and detect patterns and/or perform tasks, such as prediction or decision-making that can approximate some aspects of intelligence.[16]
- European High Level Expert Group on AI
  - » **Machine Learning:** Machine learning refers to the ability of software and computers to learn from their environments or from very large sets of representative data. This enables systems to adapt their behaviour to changing circumstances or to perform tasks for which they have not been explicitly programmed.[17]

Fundamental human rights

- ITI definition
  - » **Fundamental human rights:** Throughout the document, we use the language "impacts to fundamental human rights" or "impacts to fundamental rights." When we use these terms, we expressly mean that a use of AI could negatively impact humans in a way that jeopardizes or undermines physical integrity, health, privacy, democracy, expression, freedom, or non-discrimination (bias). Not every AI application affects fundamental rights, but those that do could more reasonably be expected to fall under stricter rules.

Explainability/Explainable AI

Explainability is a nascent field and research is being undertaken around the world to explore how explainability can provide an understanding of the path a system took to reach a particular decision or outcome. Refer to our comments on the National Institute of Standards and Technology's *Four Principles of Explainable AI* for a more in-depth explanation of ITI's views on explainability and the roles it can play in illuminating decisions to different stakeholders.[18]

- DARPA definition
  - » **Explainability:** The ability of machines to 1) explain their rationale; 2) characterize the strengths and weaknesses of their decision-making process; and 3) convey a sense of how they will operate in the future.[19]
    - In the context of explainability it may be helpful to decipher between interpretability and understandability.
      - o *Interpretability* allows a technical expert, such as an AI/machine learning expert, to understand why an algorithm made a given decision.
      - o *Understandability* enables a non-technical person (e.g., business executive or customer) to gain insight into how an algorithm works, and why it made a given decision.

References

[1] https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf.

[2] Refer to the section exploring potential AI use cases for more on this topic.

[3] See glossary for a definition of key terms, including explainability and transparency.

[4] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[5] See definition of fundamental human rights in the glossary.

[6] For example, the GDPR recognizes that ensuring network and information security is a "legitimate interest" of entities for processing personal data (Recital 49).

[7] OECD Principles on Artificial Intelligence, available here: https://www.oecd.org/going-digital/ai/principles/

[8] GPAI information available here: https://gpai.ai/about/

[9] More about AI in financial services available here: https://finreglab.org/faqs-about-ai-in-financial-services

[10] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[11] Definition from 2017 ITI AI Policy Principles.

[12] https://www.nap.edu/catalog/25021/the-frontiers-of-machine-learning-2017-raymond-and-beverly-sackler

[13] https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

[14] https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[15] https://www.weforum.org/agenda/2018/11/what-is-adversarial-artificial-intelligence-is-and-why-does-it-matter/#:~:text=Adversarial%20AI%20is%20the%20malicious,discover%20meaning%20from%20complex%20data.

[16] *Ian Goodfellow Yoshua Bengio Aaron Courville, Deep Learning (Adaptive Computation and Machine Learning series), (MIT Press, 2016), 1.*

[17] https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

[18] https://www.itic.org/policy/ITICommentsNISTIR8312ExplainableAI.pdf

[19] Explainable AI (XAI), information here: https://www.darpa.mil/program/explainable-artificial-intelligence

ITI

Promoting Innovation Worldwide

www.itic.org