

Leveraging AI for Global Prosperity by Accelerating the UN Sustainable Development Goals Achievement



Prof. Mohamed Essaïdi

EMSI, Morocco

IEEE Special Interest Group on Humanitarian Technologies, Global Chair

m.essaïdi@emsi.ma | essaïdi@ieee.org



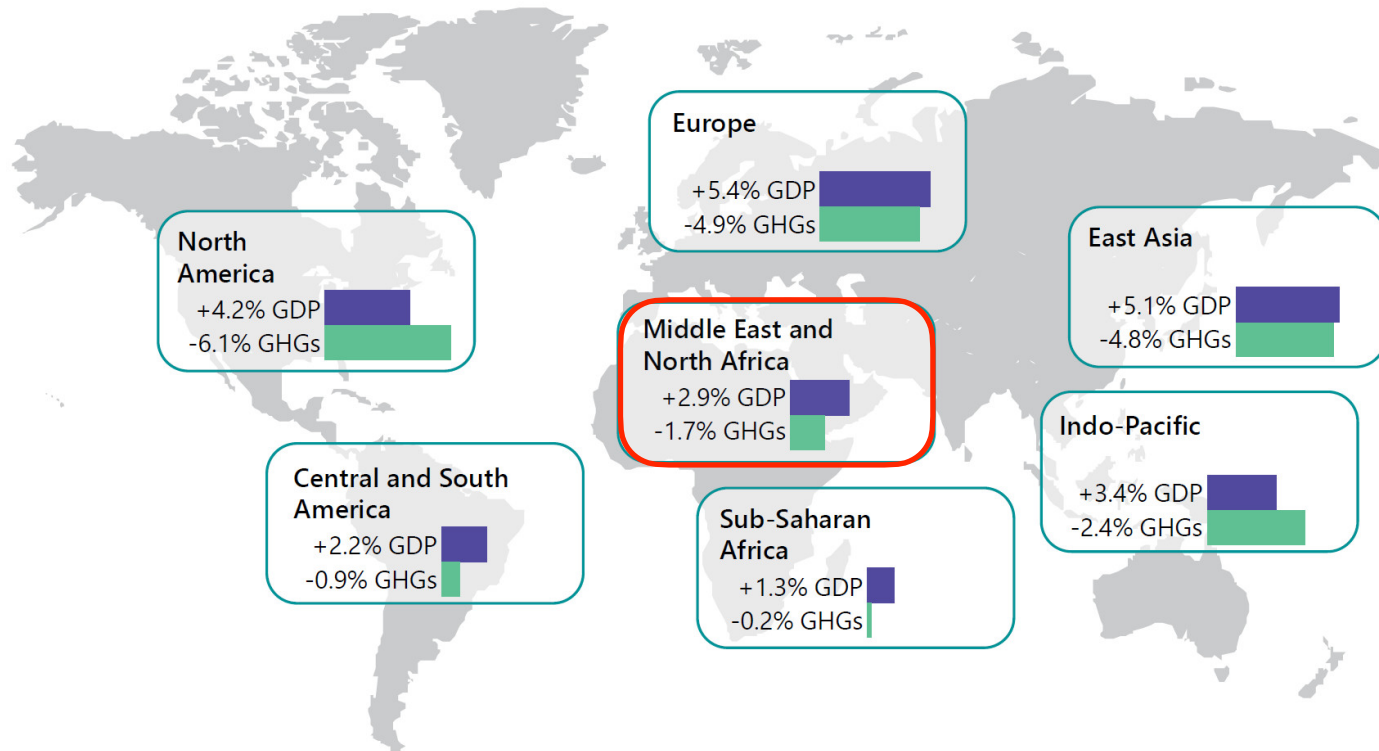
9th International Conference on Digital Economy Emerging Technologies and Business Innovation

FSJES Souissi, UM5, Rabat, Morocco | May 16, 2024

Outline

- Introduction
- UN Sustainable Development Agenda
- AI for SDGs
- Key AI initiatives for SDGs
- Morocco's NDM vs UN SDGs
- AI for NDM acceleration
- Concluding Remarks

AI will contribute \$15.7 trillion to global economy by 2030



Source: PwC analysis

Using AI for environmental applications could:

Contribute

\$5.2
TRILLION USD

to the global economy
in 2030, a 4.4% increase
relative to business as usual

Create

38.2
MILLION

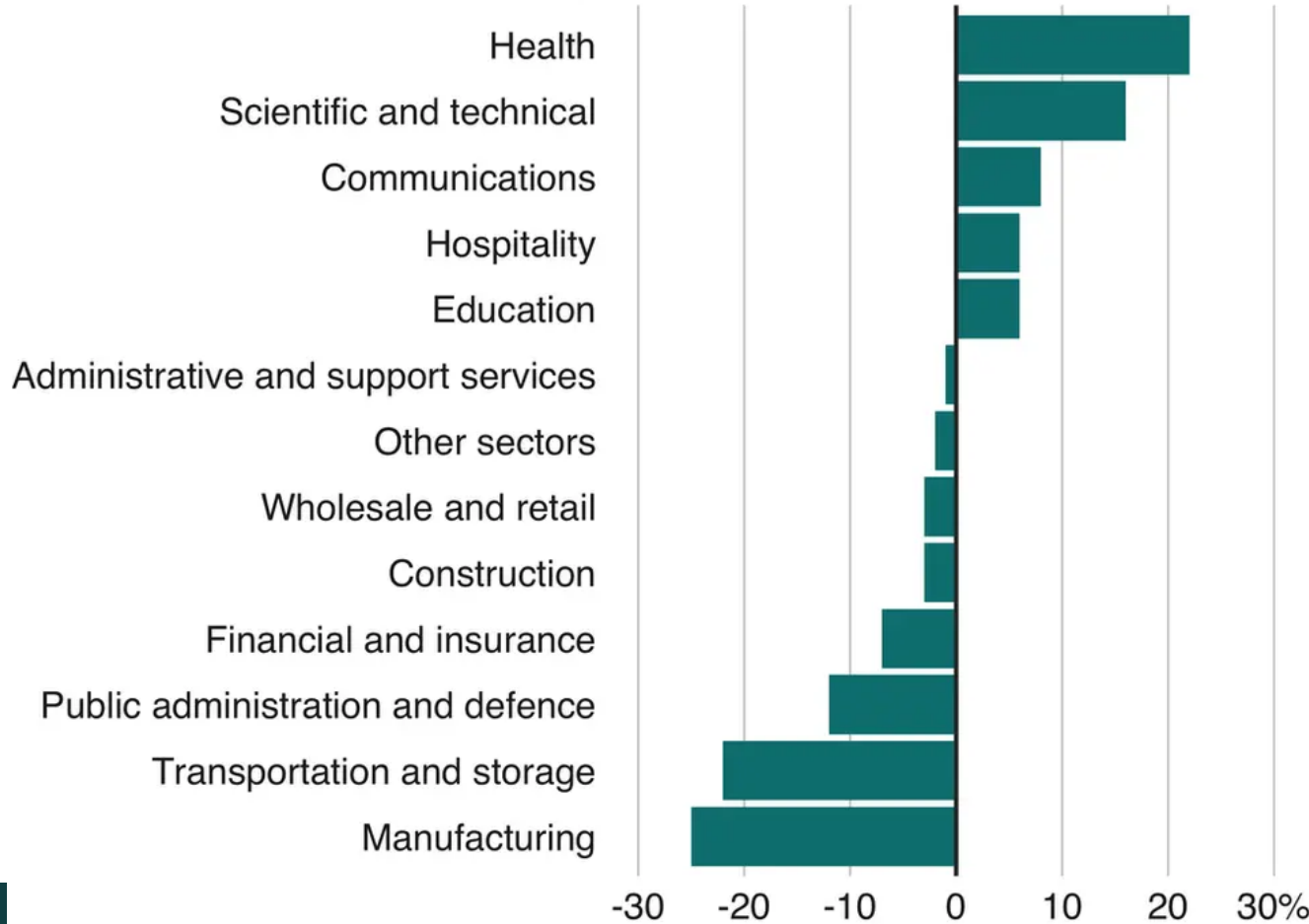
net new jobs across
the global economy

Reduce worldwide GHG emissions by 4% in 2030,
an amount equivalent to

2.4 GT CO2E –

How AI could change the job market

Estimated net job creation by industry sector, 2017-2037



Source: PwC

BBC

What jobs will Gen AI create?



Large language models (LLMs) will transform collaboration between humans and AI, reshaping job roles. While outcomes remain uncertain, potential new job areas could emerge with LLM adoption.



AI Model and Prompt Engineers



Interface and Interaction Designers



AI Content Creators



Data Curators and Trainers



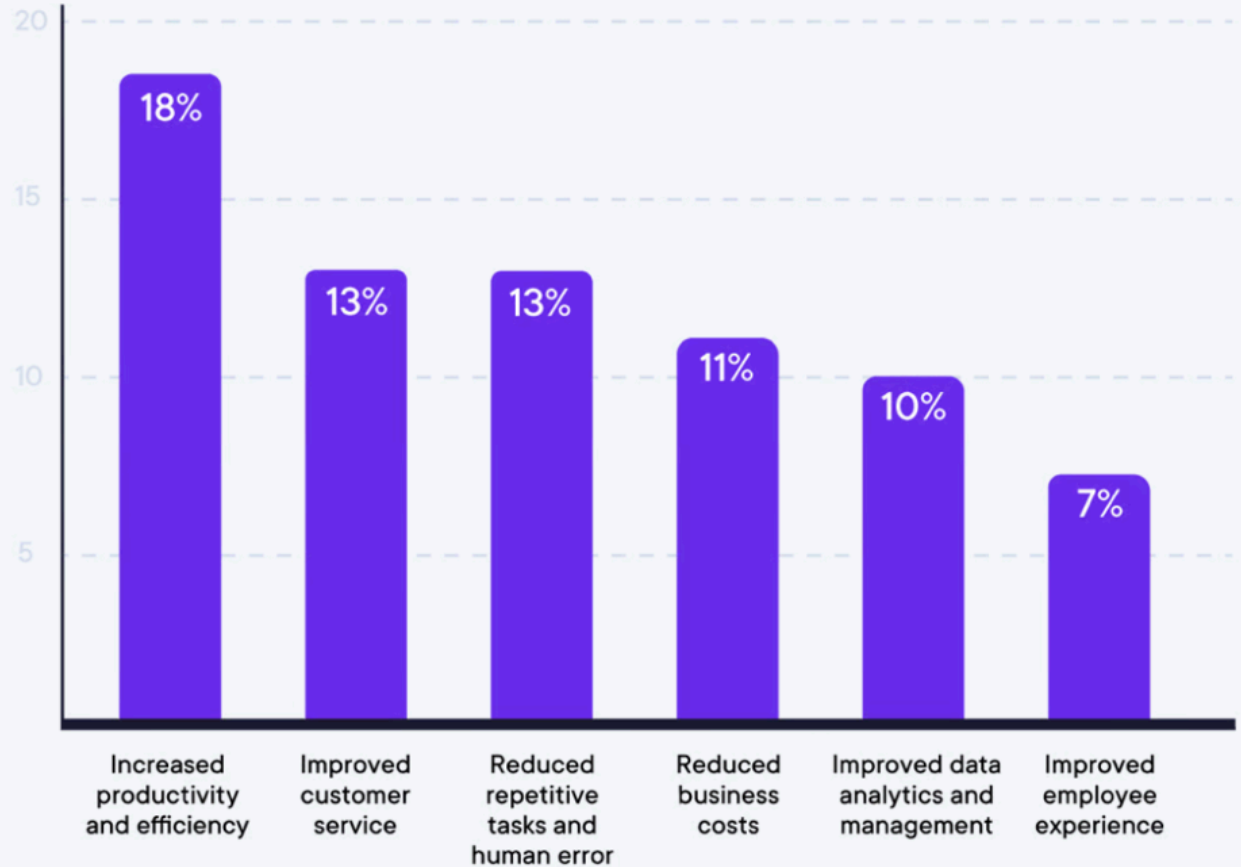
Ethics and Governance Specialists

Benefits of using AI for organizations

2024

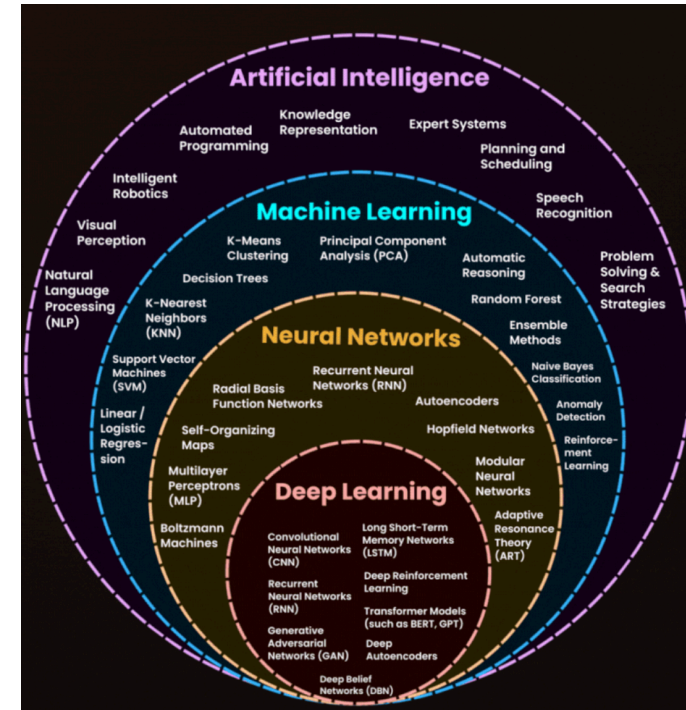
Pluralsight AI skills report

The gap between AI investments
and worker readiness



What is Artificial Intelligence ?

- **Artificial Intelligence (IA)** can be defined as the ability of software systems to carry out tasks that usually require human intelligence, such as vision, speech, language, knowledge, and research.
- **Machine Learning (ML)** is the ability of computer algorithms to learn from data and improve automatically.
- **Natural language processing (NLP)** is a machine learning technology that gives computers the ability to interpret, manipulate, and comprehend human language.
- **Computer vision** is a field of AI that ML and neural networks to teach computers and systems to derive meaningful information from images, videos and other visual inputs.
- **Artificial Neural Networks (ANN)** are artificial intelligence algorithms that learn relationships between different data sets in a manner similar to how the human brain analyzes this information.



Gartner Top 10 Strategic Technology Trends **2024**

1. AI as Partner: AI Trust, Risk and Security Management (AI TRiSM)
2. Be Safe: Continuous Threat Exposure Management (CTEM)
3. Protect the Future: Sustainable Technology
4. Developer-Driven Self-Service: Platform Engineering
5. Accelerate Creation: AI-Augmented Development
6. Tailor Your Tailor's Work: Industry Cloud Platforms
7. Optimize Decision-Making: Intelligent Applications
8. Power AND Responsibility: Democratized Generative AI
9. Push the Pioneers: Augmented Connected Workforce
10. Buyers With Byte(s): Machine Customers

Technology Predictions for 2024

Sorted by Tech. Development

1. **Generative AI Applications (A/B):** Generative AI use will increase with rapidly expanding efficiency and new applications and services both beneficial and detrimental. Ethical and societal issues will continue to rise. Expect strong short-term impacts on business, education and society.
2. **Next Generation AI (B+):** The evolving advancements and developments in the field of artificial intelligence that push the boundaries beyond current capabilities. It is the next generation of Artificial Intelligence (AI) that is expected to be more advanced and sophisticated than the current AI systems.
3. **Advances in Cybersecurity (B)** will enhance public confidence and will enable reliance on the cyber infrastructure for large scale applications including energy production and distribution.
4. **Managing Misinformation (B):** AI deepfakes (text, audio, visual) will become regular tools that will require careful management.
5. **Remote Healthcare (B),** monitoring sensors and system-level data integration will enable patients to obtain remote medical assistance, physicians to improve diagnosis and treatment, optimal utilization of individuals' medical history, and efficient health care delivery protocols.
6. **Digital Twins for Vertical Applications (B)** will advance state of the art of predictions, what-if-analysis and oversight in a number of industries, such as data centers, medicine, geo-physical hazards, manufacturing, agriculture, transportation, and many others.
7. **New 3D Printing Applications (B)** will evolve towards customized and automated solutions in many domains.
8. **New Programming Models (B-).** Advances in AI, broader adoption of script-based languages, and further digital transformation into non-programmers' world will further increase ease of development and require new programming models and DevOps, such as serverless, from the Edge to Cloud.
9. **Reliability (B-)** will emerge as a major concern in a widespread set of application fields.
10. **Autonomic Autonomous and Hybrid Systems (B-)** will see increased development and adoption in areas, such as driving, laboratory work, agriculture, and many others.
11. **Distributed Energy Resources for Powering Data Centers (B-)** Engaging renewable energy based on distributed energy resources for powering data centers will have a high impact on clean energy requirements for data centers.
12. **Sustainable ICT (B-):** will evolve by designing, manufacturing, using, and disposing of electronic systems efficiently and effectively for new use cases, with minimal or no impact on the environment.
13. **Regenerative AgriTech (B/C)** is a holistic, circular approach to farming that strives to improve the health of agroecosystems and the natural ecosystems that support them.
14. **Non-Terrestrial-Networks (B/C)** involving satellites and high-altitude platforms (HAPs) expand and augment the capabilities of terrestrial networks (TN) involving wireless and cabled communications in the quest to connect everything to everything (E2E) in real time (RT).
15. **New Battery Chemistry and Architecture (B/C)** will replace Lithium and will make it possible to make batteries that are cheaper and more sustainable.
16. **Low Power AI Accelerators (B/C)** will be key-components for practical, compact, cost-effective, long-term reliable computation units for Self-driving vehicles and AI robots, data-centers, LLM, systems, smart phones, games.
17. **Alternate Materials for Electro Machines (EV motors) (C+):** Inadequate raw materials for conventional high-performance electro machines motivates discovery and engineering.
18. **Cost Effective Recycling of Batteries (e.g. Lithium) (C+)** to recover materials for reuse will reduce the need for mining and increase the general sustainability of battery technology.
19. **Metaverse (C+)** will bridge the gap between the real and the digital worlds, by solving real world industrial problems digitally.
20. **Accessible Quantum Computing (C-)** will improve public understanding and access to the power of quantum computing, increasing 'conventional' computing efficacy exponentially.
21. **Satellite (Constellation) Recycling (C/D)** will enable circular economy in space ensuring long term sustainability. We expect an initial success in 2024 with increasing awareness of the tremendous impact on Humanity.

Key milestones

1665 The first scientific journal is printed. ¹	1768 Encyclopedia Britannica publishes its first edition. ²	1873 The Dewey Decimal System is developed for Amherst College Library. ³	1967 The “Ask NYPL” (New York Public Library) hotline opens. ⁴	1967 ORBIT launches as a database search service for research abstracts. ⁵	1975 Ohio State University implements the first major digital catalog. ⁶	1990 Three McGill University students build Archie, the first search engine. ⁷
1996 Ask Jeeves is founded. ⁸	1998 Google goes online with its PageRank algorithm. ⁹	2001 Wikipedia launches. ¹⁰	2008 Stack Overflow begins crowdsourcing programming questions and answers. ¹¹	2010 Microsoft introduces SharePoint for enterprises. ¹²	2012 Google announces its knowledge graph, a significant step toward semantic search. ¹³	2019 Researchers propose K-BERT, a knowledge graph-enabled LLM. ¹⁴
2022 OpenAI releases ChatGPT. ¹⁵	2023 Bing Chat is unveiled by Microsoft. ¹⁶	2025 A leading airline will announce that customers are just as satisfied with chatbot agents as human agents.	2027 Data poisoning (adding malicious data to ML models) will be a top cybersecurity threat to enterprises.	2028 Major corporations will have proprietary chatbots to assist with knowledge management, research, and task completion.	2029 AI advisors will receive more search traffic than traditional search engines.	2031 A smartphone will launch that replaces the app-based interface with an agent-based one.

Strategic and Responsible Use of Artificial Intelligence in the Public Sector

Improving government efficiency and decision making

Relationships with and services for citizens and businesses

Public safety and security

Regulatory functions

Healthcare

Transportation

Sustainable development goals (SDGs)

Public integrity and accountability

Education

OECD Public Governance Reviews



**The Strategic and Responsible
Use of Artificial Intelligence
in the Public Sector of Latin
America and the Caribbean**



March 2022

Five emerging AI use cases in GPS



Spotting trouble before it occurs

Video Surveillance Predictions

Using AI and computer vision-enabled video surveillance to detect potential security threats more quickly and accurately.



The art of war in the AI era

Agent-based Simulations to Refine Military Strategy

Using deep learning to simulate tactical moves and refine military strategy in real time.



City of the future

Civil Asset and Infrastructure Management

Using AI to monitor and maintain a city's physical assets and infrastructure, ensuring they are fully functional and operating safely.



Augmenting and assisting the judgement of our judges

Legal Outcome Predictions

Using machine learning and deep learning to analyze decades of case law — and millions of past cases — to predict outcomes for future cases and accelerate case resolutions in both domestic and international courts.



Making adaptive learning truly adaptive

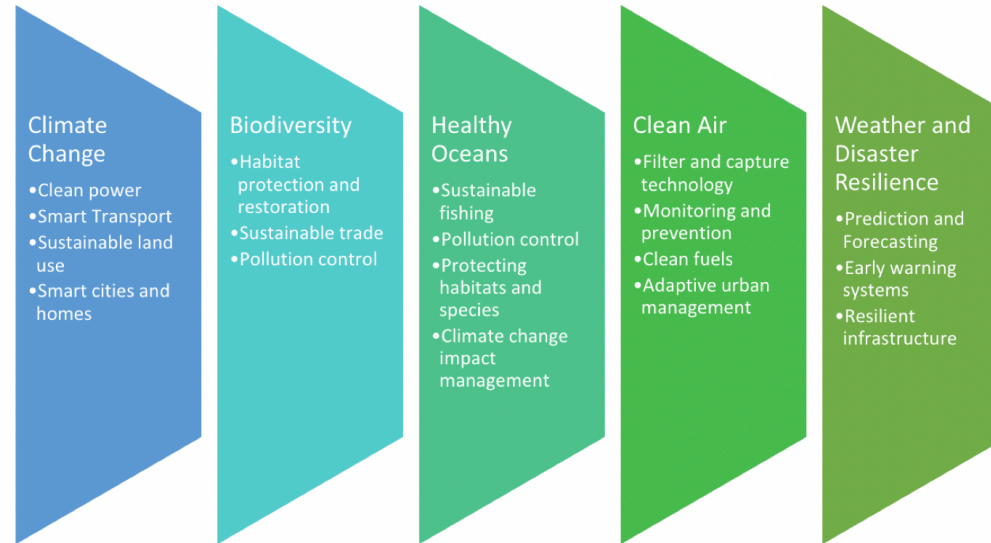
Education Tech: Learning Analytics for Adaptive Learning

Using AI to deliver a one-on-one education experience that truly adapts to the needs and abilities of the learner.

GPS: Government & Public Services

AI & Sustainable Development

- Advancements in AI present new approaches for sustainable development
- Potential applications of AI technologies for environmental sustainability highlighted by World Economic Forum in 2018



Source: Adapted from WEF (2018)

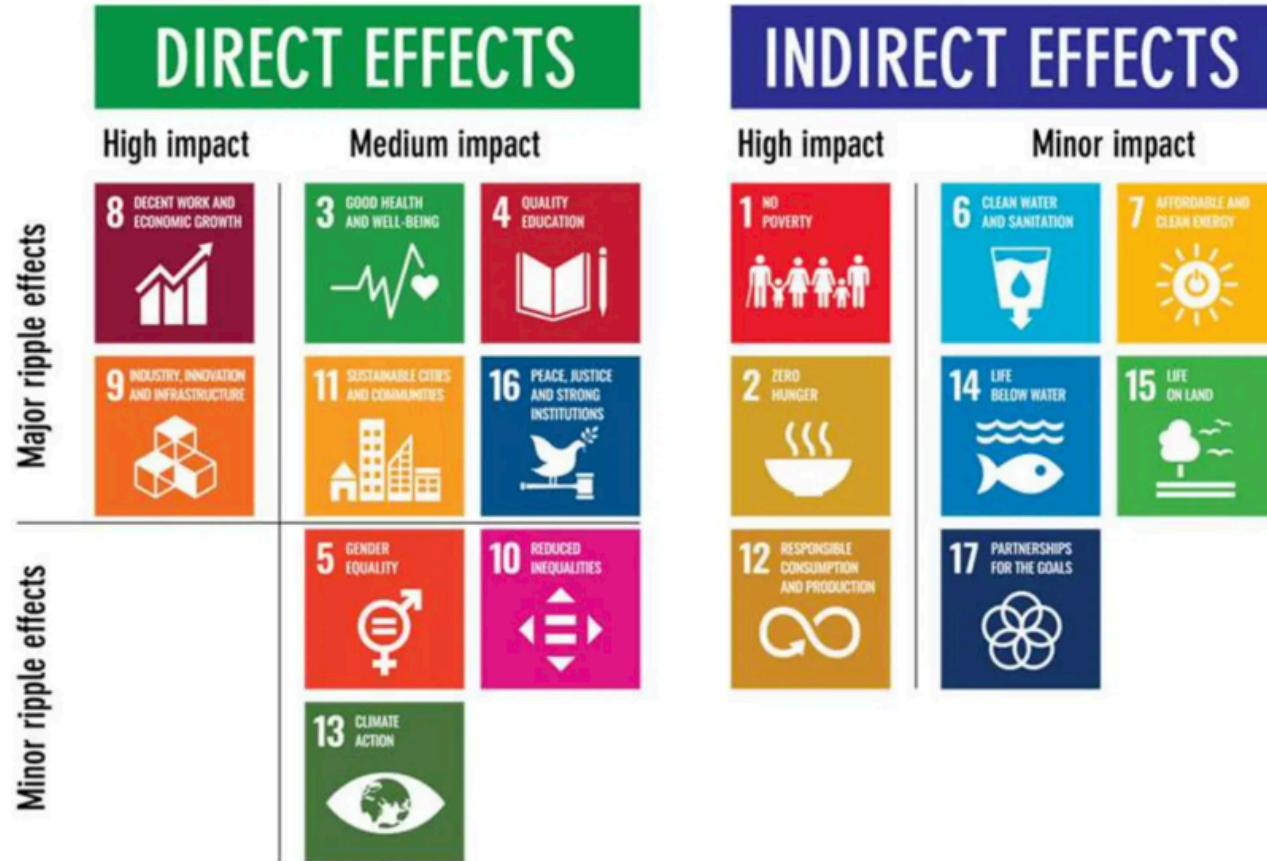
UN Sustainable Development Agenda (SDGs)



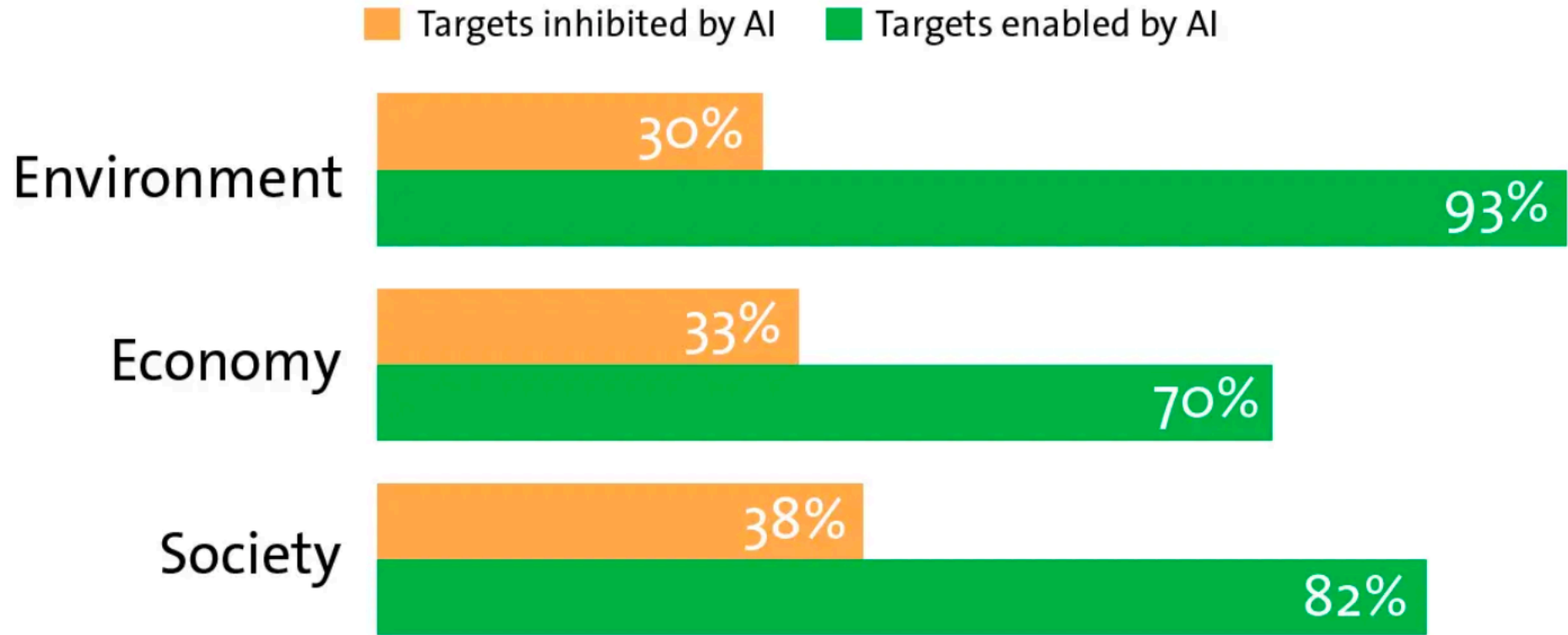
Other Categorization of the SDGs



Categorizing The SDGs in terms of AI Impact

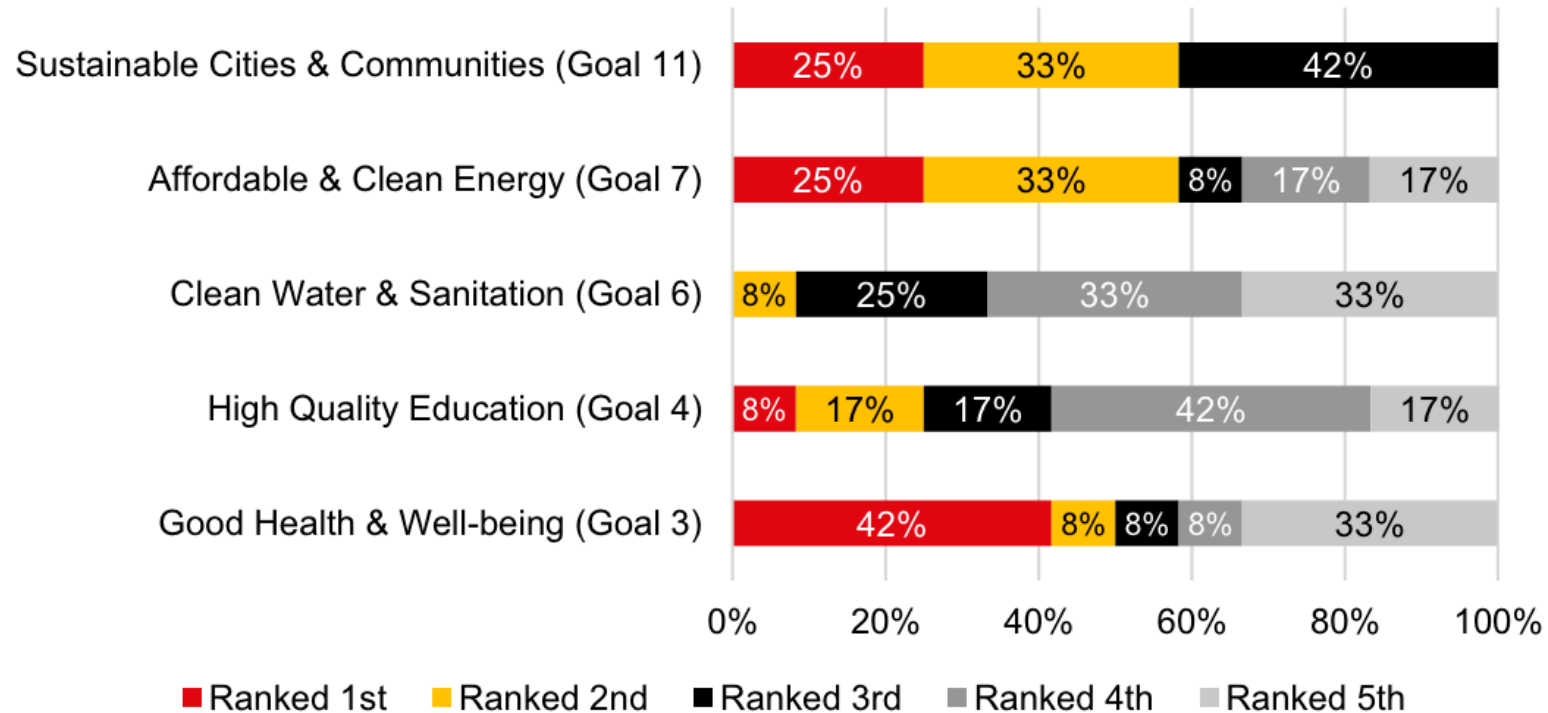


Impact of AI on the achievement of the SDGs

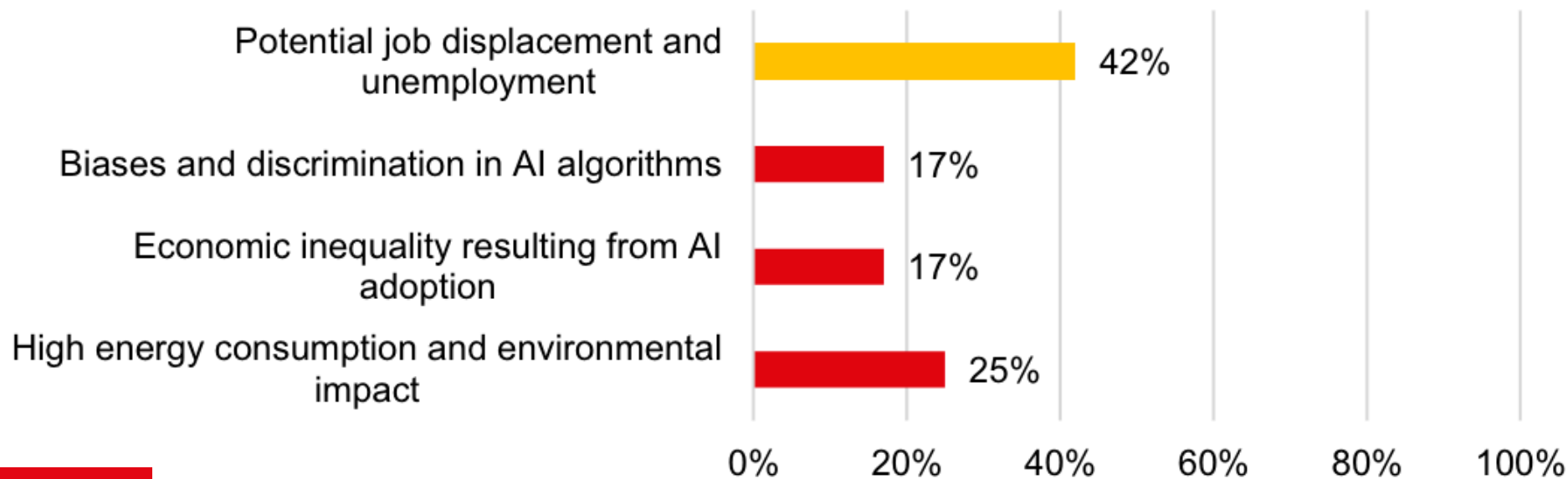


Source: Springer Nature Sustainability Community.

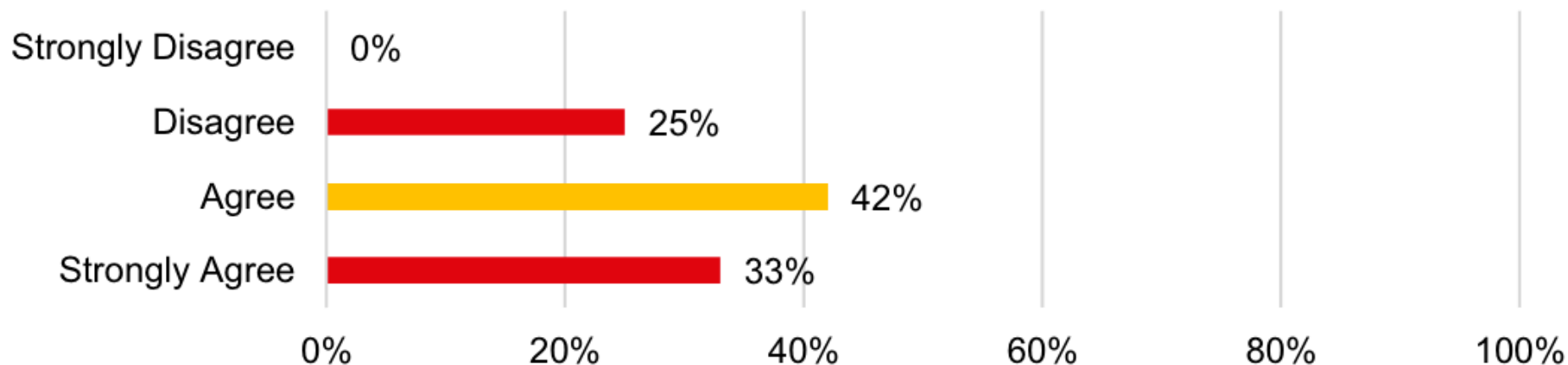
Which of the following UN SDGs stands to gain the most from the implementation of AI?



Which challenge arising from AI poses the greatest obstacle to achieving the UN Sustainable Development Goals?



The advantages of using AI to realise the UN SDGs outweigh the potential challenges and risks.



Which of the following measures should be prioritised for ensuring the responsible and sustainable development of AI?



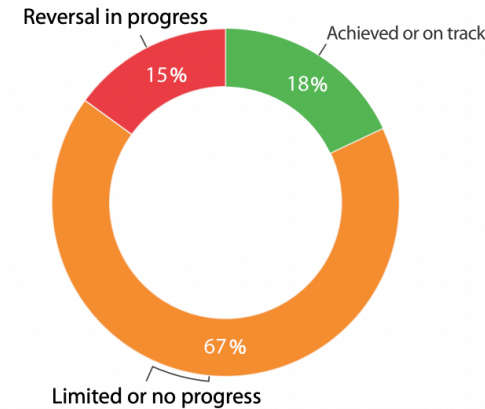
Sustainable Development Report 2023



The SDG Index and Dashboards



■ Major challenges
 ■ Significant challenges
 ■ Challenges remain
 ■ SDG achieved
↓ Decreasing
 → Stagnating
 ↗ Moderately improving
 ↑ On track or maintaining SDG achievement



Source: Authors' analysis

The world SDG dashboard at the midpoint of the 2030 Agenda



SUIVI DES OBJECTIFS DE DEVELOPPEMENT DURABLE



Artificial intelligence to accelerate the SDGs

Artificial intelligence (AI) is a disruptive technology that can revolutionize society and an enabling technology that can foster societal progress

AI development should be used to promote the achievement of the global SDGs



AI research organizations, universities, and companies all have a responsibility to use AI to enhance the long-term growth of society, economy, and environment

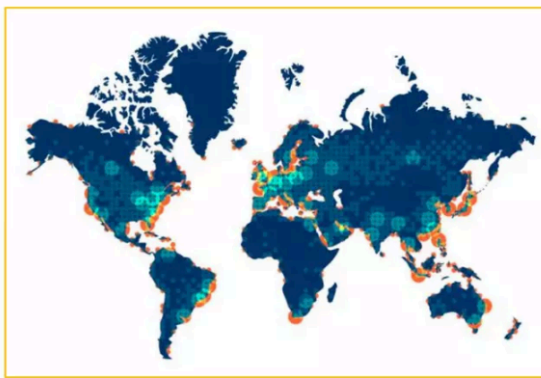
The global effort to AI to assist sustainable development, human rights, and whole of humanity, leaving no one behind

Society 5.0: Japan AI Enabled Development Model



Japan Development Model

Source: Society 5.0, <https://www8.cao.go.jp/>



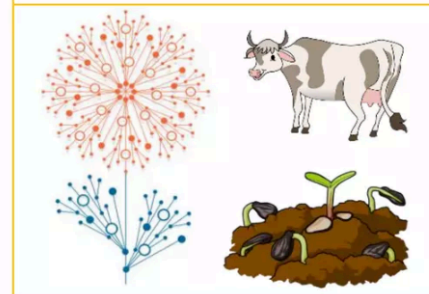
Satellite images analytics and Agro-meteorological monitoring



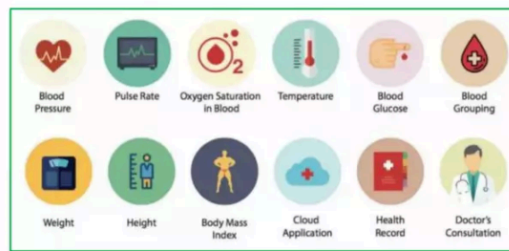
Predict crop/animal diseases, pesticide planning, livestock management, etc.



In-field monitoring, harvesting and picking, weed control, autonomous mowing, sorting and packing, etc.



Assessing Soil/Crop/Livestock Health



Preventive Healthcare

(Discover measures to disease prevention, as opposed to disease treatment)

3 GOOD HEALTH AND WELL-BEING



Predictive Healthcare

(analyzes historical data to prevent future target events)



Personalized Healthcare

(Right patient, right drug, right dose and the right time)



Cognitive Healthcare

(automates decisions using human-like analysis)

coursera UDACITY



- **Interactive Massive Open Online Courses (MOOC)**
- **Intelligent tutoring system (ITS)**
- **Machine Teaching**

4 QUALITY EDUCATION



kaggle

HackerRank



CodinGame

CrowdANALYTIX



[topcoder]

ImagineCup



Challenge-based Learning (CBL)



Flexible Content and Tools

Instructional materials allow for differentiated path, pace, and performance tasks



Targeted Instruction

Instruction aligns to specific student needs and learning goals



Student Reflection and Ownership

Ongoing student reflection promotes ownership of learning



Data Driven Decisions

Frequent data collection informs instructional decisions and groupings

Personalized Learning

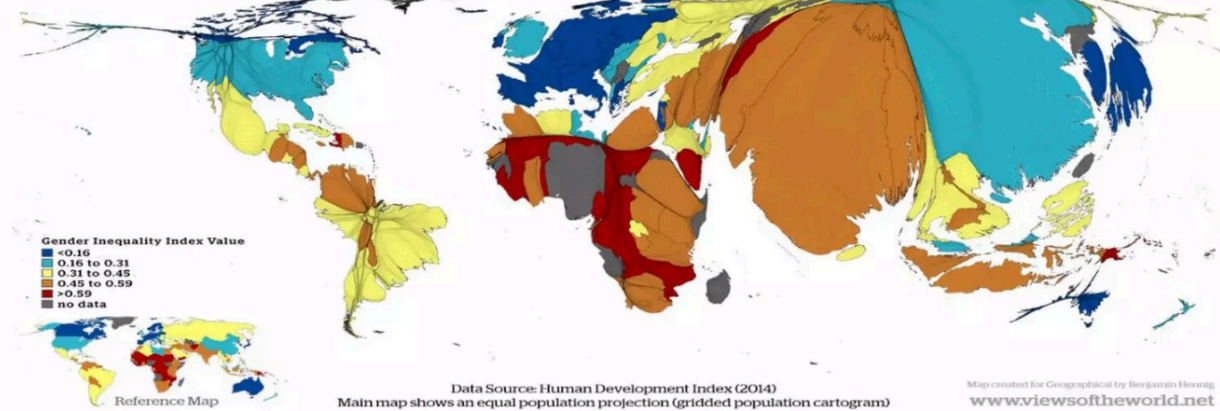


Global Classroom
(Mixed reality, computer vision and monitoring performance)

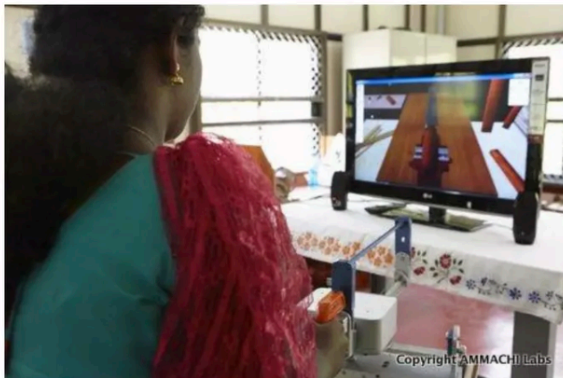


- Monitoring and track **gender bias** and provide actionable insights to the decision makers to drive balanced hiring

Gender Inequality



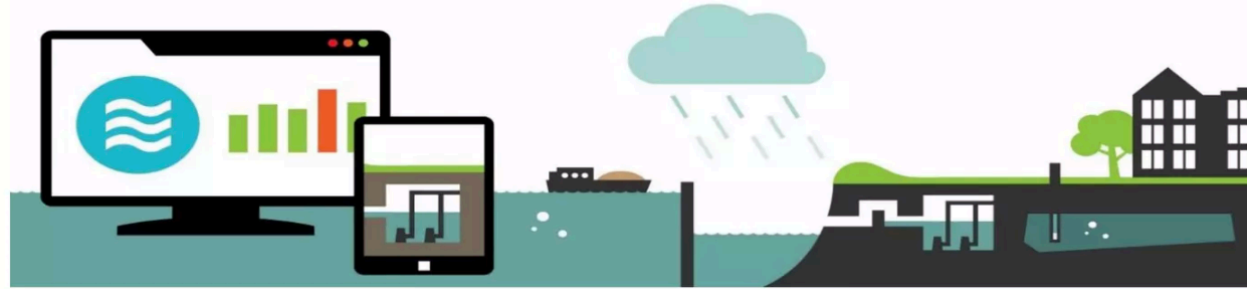
- Women economic empowerment** (innovative technologies for skill development and training women in vocational trades)



6 CLEAN WATER AND SANITATION



- **Smart water monitoring and management systems**

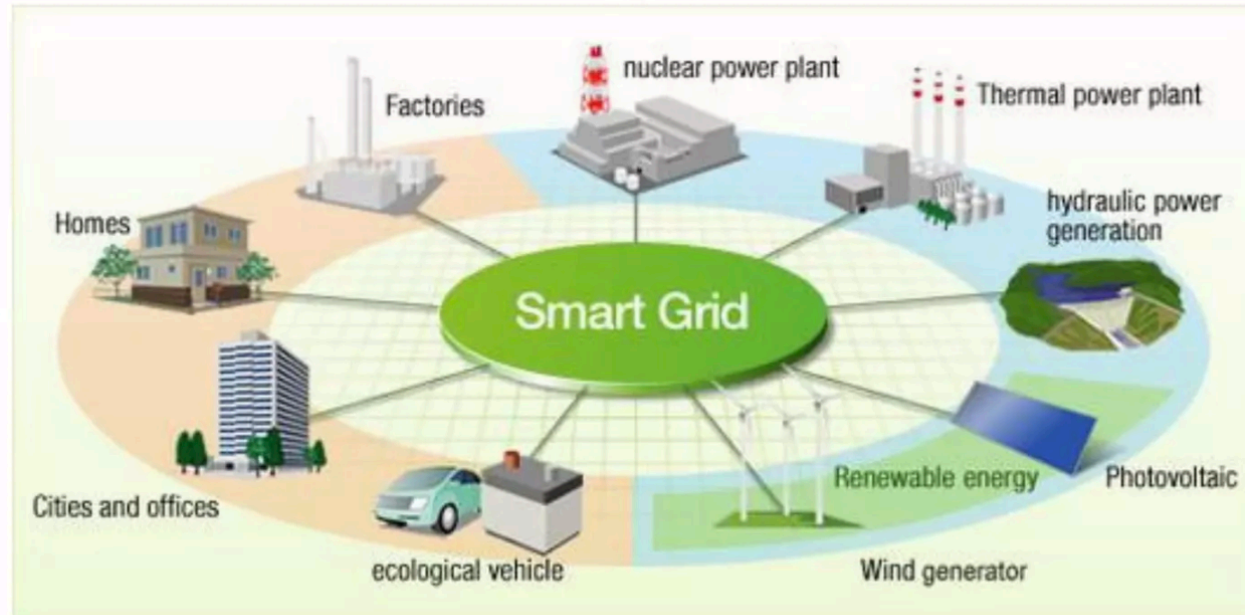


wateralliance

7 AFFORDABLE AND CLEAN ENERGY



- **Smart renewable energy grids**



Better World
SOLUTIONS

Improve Job Security (Utilizing AI to Automate Mundane Tasks)



8 DECENT WORK AND
ECONOMIC GROWTH

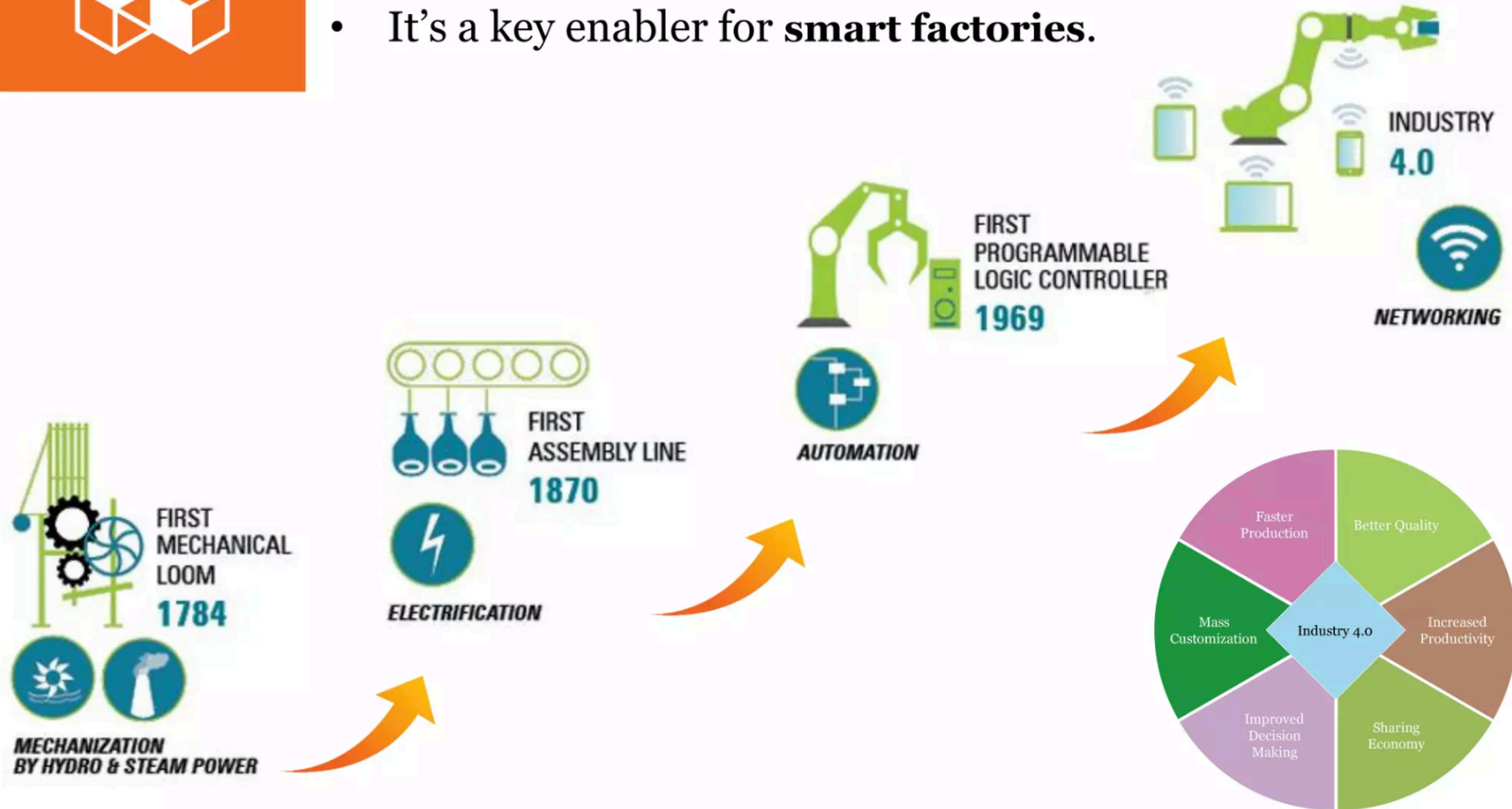


AI-Powered Training Solutions



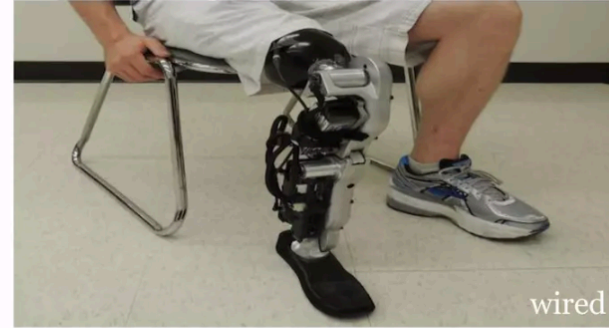


- AI plays instrumental role in the 4th industry revolution **Industry 4.0** that relies on digitalization, automation, connectivity and analytics.
- It's a key enabler for **smart factories**.





- Robotic assistive systems will empower **elderly and physically challenged** leading to more equal and inclusive society.



11 SUSTAINABLE CITIES AND COMMUNITIES



- Multimodal **smart sensors** to monitor different aspects in the city/community and **manage assets and resources** efficiently
- **Smart cities** to bring in efficiency and bottom-line benefits as well as environmental improvements.

12 RESPONSIBLE CONSUMPTION AND PRODUCTION



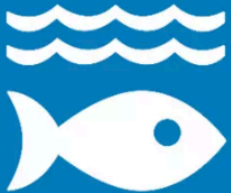
- Monitor **consumption levels**
- Predict **optimal production levels** to reduce waste.

13 CLIMATE ACTION



- Global Earth Observation System of Systems
- Track stratospheric ozone depletion
- Model **climate change** to **predict disasters** such as windstorms

14 LIFE BELOW WATER



- Track **illegal fishing activities** through pattern recognition
- Track **marine-life migration**
- Underwater **exploration** using submarine robots (example: Stanford OceanOne).

15 LIFE ON LAND



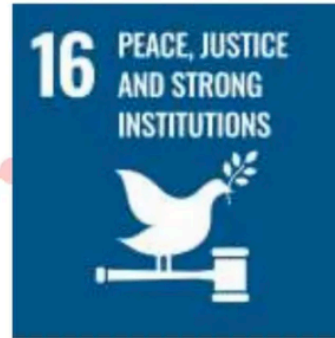
- Monitor airborne, marine, water pollutants, **species health, land-use change**, food security and nutrition, noise levels, weather-related stresses and **disease vectors**.
- Predict **population trends, desertification trends, epidemics**, etc.

16 PEACE, JUSTICE AND STRONG INSTITUTIONS



- Better **surveillance**, reconnaissance, **demining**, counter-IEDs, planning and decision making technologies
- Enable efficiencies, **transparency**, engagement, **personalized and responsive intelligent services** via Smart Governance.

Reduce the amount of time and resources required for data gathering, analysis, and option creation, and allocate those resources **instead to more challenging tasks like strategic decision-making, dialogue, negotiation, and trust-building**



Algorithms driven by AI can sift through enormous volumes of data to find possible partners that share their values, objectives, and vision

Global AI Strategy Landscape

50 National Artificial Intelligence Policies as at February 2020.

 Argentina Drafting the "National Plan of Artificial Intelligence". Falls under the Innovative Argentina 2030 Plan and the 2030 Digital Agenda.	 Australia November 2019, AI Roadmap focused on specialization in health, infrastructure and natural resources. Planning for an additional 16,000 AI specialists by 2030.	 Austria June 2019, "Artificial Intelligence Mission Austria 2030 (AIM AT 2030)". Outlines seven fields for which AI will be critical.	 Belgium March 2019, "AI 4 Belgium" launched and includes seven major objectives.	 Brazil Consultation period ended January 2020. Building a network of eight research facilities focused on artificial intelligence.
 Canada 2017 federal budget announced five-year, \$125m plan. Led by CIFAR Research and talent focus. First National AI Strategy.	 Chile Expected April 2020. Ministry of Science, Technology, Knowledge, and Innovation created a committee of 10 experts to develop.	 China July 2017, China launched the most comprehensive AI strategy globally with 2030 targets for a \$1T RMB AI industry.	 Colombia November 2019, first draft issued for "National Policy for Digital Transformation". Medellín to become an AI & Robotics Centre of Excellence.	 Czech Republic May 2019, "National Artificial Intelligence Strategy of the Czech Republic" was launched.
 Denmark March 2019, Denmark announced the "National Strategy for Artificial Intelligence" with four key objectives.	 Estonia - Kratts Strategy May 2019, Estonian AI experts, led by government CIO produced a roadmap, later adopted as the Estonian National AI Strategy in July 2019.	 Finland June 2019, "Leading the Way into the Age of Artificial Intelligence" identified 11 key actions following May 2017 Steering Group announcement.	 France €15 billion plan announced in 2018 influenced by the "Villani Report" to transform France into a global leader in AI.	 Germany €3 billion plan announced Nov 2018 with a dedicated AI strategy to make Germany & Europe a global leader in AI.
 Hungary October 2019, Hungary announced an AI Action Plan, the first pillar of a national AI strategy, expected in 2020.	 India June 2018 working paper on using AI to ensure social growth, inclusion and positioning the country as a leader in AI.	 Indonesia Indonesia Artificial Intelligence Society (IAIS) inaugurated under Smart Indonesia in October 2019. National Strategy expected in 2020.	 Ireland Irish Economic Development Agency led process. AI Master program launched in 2018 and is 100% industry driven.	 Israel Innovation Authority, tasked with AI policies, has warned that a strategy is needed to prevent falling behind.
 Italy March 2017, AGID released a White Paper called "AI at the service of citizens," which was edited by the AI Task Force.	 Japan March 2017, Japan's AI policy, the "Artificial Intelligence Technology Strategy", was announced second only to Canada with "Society 5.0".	 Kenya January 2018, government announced task force to create a five-year strategy on national use of emerging technologies.	 Lithuania April 2019, Artificial Intelligence Strategy announced "to modernize and expand the current AI ecosystem and ensure that the nation is ready".	 Luxembourg May 2019, launched "Artificial Intelligence, a strategic vision for Luxembourg".
 Malaysia 2018, Malaysia revealed a National Artificial Intelligence Framework expanding the National Big Data Analytics Framework.	 Malta October 2019, "A Strategy and Vision for Artificial Intelligence in Malta 2030" Malta.ai launched and aspiring to be the "Ultimate AI Launchpad".	 Mexico June 2018, "Towards an AI Strategy in Mexico: Harnessing the AI Revolution", serves as a foundation for building full AI strategy.	 Netherlands November 2018, AINED published a roadmap for developing a full national strategy.	 New Zealand May 2018, AI Forum of New Zealand, released "Artificial Intelligence: Shaping a Future New Zealand".
 Norway January 2020, Norway issued its National Strategy for Artificial Intelligence.	 Pakistan Presidential Initiative for Artificial Intelligence launched December 2018, focused on training beginners in AI and advanced technology.	 Philippines Nov 2019, AIM, Abaitiz School of Innovation, Technology and Entrepreneurship (ASITE) appointed to craft an AI roadmap.	 Poland November 2019, "Assumptions for the AI strategy in Poland" as an action plan towards developing an AI strategy.	 Portugal February 2019, "AI Portugal 2030", seeks strengthen economic growth, scientific excellence, and human development using with AI.
 Qatar October 2019, National AI Strategy as a blueprint produced by Qatar Computing Research Institute (QCRI).	 Russia October 2019, Russia published its National Strategy for the Development of Artificial Intelligence by 2030.	 Saudi Arabia September 2019, Royal decree to establish an AI center, to align with the Kingdom's Vision 2030 program.	 Singapore May 2017, AI Singapore is a five-year, \$150 million national program launched in to enhance Singapore's capabilities in AI.	 South Africa Intsimbi Future Production Technologies Initiative" launched in 2018 with aim to advancing manufacturing sector.
 South Korea May 2018, five-year AI development plan launched with \$195B budget.	 Spain March 2019, the Spanish Ministry of Science, Innovation and Universities launched the RDI Strategy in Artificial Intelligence.	 Sweden National Approach for Artificial Intelligence launched in May 2018.	 Switzerland An Artificial Intelligence (AI) expert group has published its recommendations for a Swiss AI strategy.	 Thailand Thailand's Digital Economy and Society (DES) Ministry has drafted the country's first artificial intelligence (AI) ethics guidelines.
 Tunisia AI Task Force and Steering Committee to develop a national AI strategy. The strategy was scheduled to be published in the first quarter of 2019.	 United Arab Emirates October 2017 announced strategy. First country to create a Ministry of AI and first in the Middle East to launch an AI strategy.	 United Kingdom April 2018, "Sector Deal" announced. \$124B funding as part of the UK's larger industrial strategy.	 United States of America February 2019 by Executive Order to promote and protect AI technology. AI.gov launched Mar 2019. Followed by the National Artificial Intelligence Research and Development Strategic Plan.	 Vietnam Ministry of Information and Communications developing a broad AI strategy.

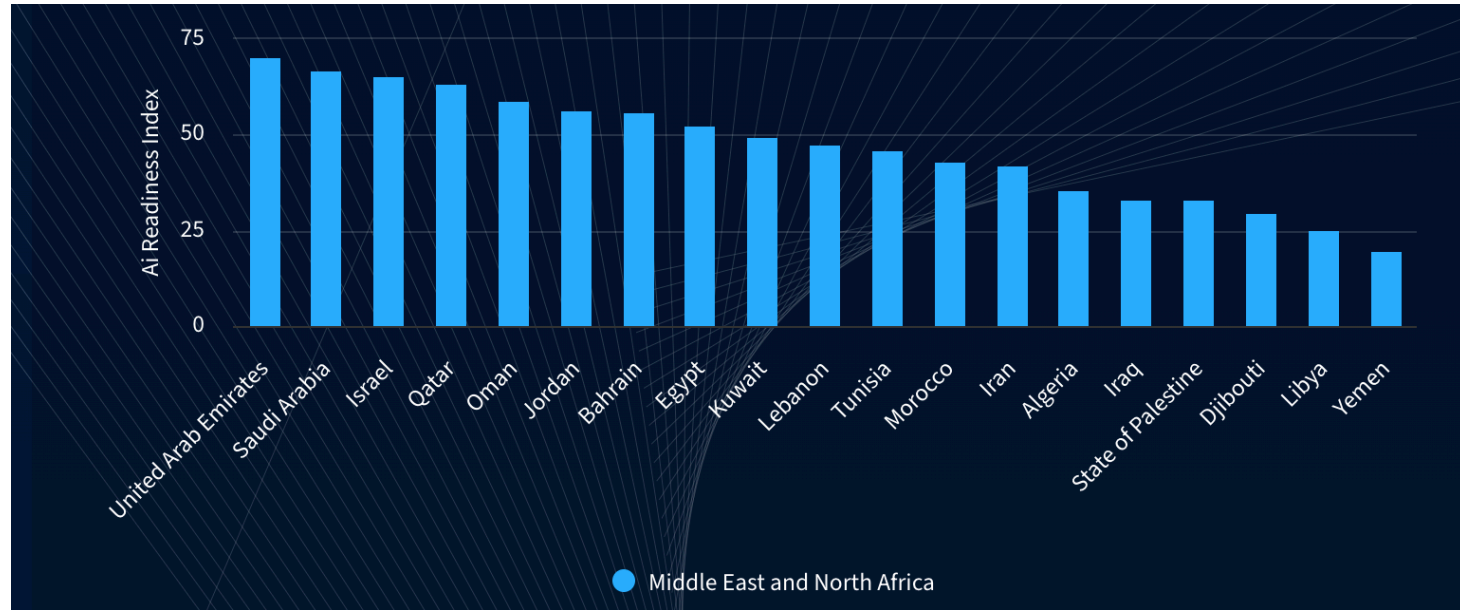
Source: HolonIQ and source government strategy and policy papers.

www.holoniq.com

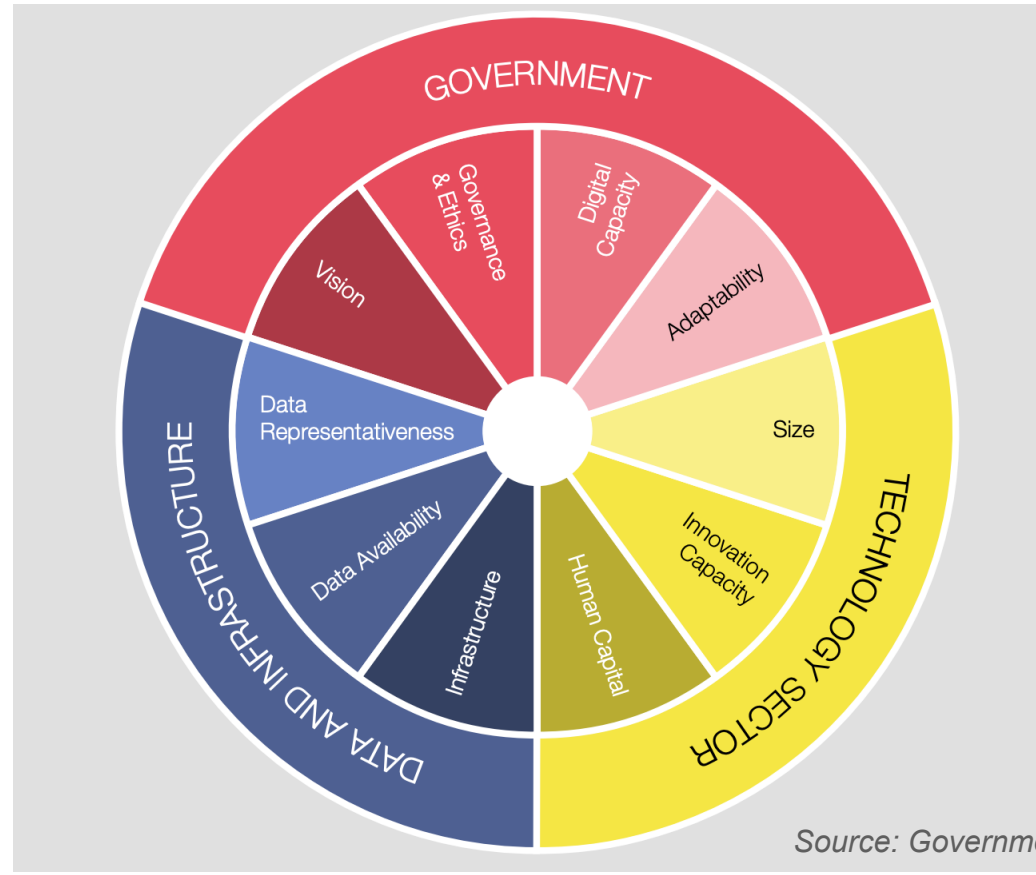
2023 Government AI Readiness Index in MEA



Government AI
Readiness Index
2023



Pillars & Dimensions of the Government AI Readiness Index



Source: Government Readiness AI Index, 2021.

AI Readiness Index Indicators (11)

Cluster	Indicator
Governance	Data protection/privacy laws - yes/no
	National AI strategy - yes/no/pending
Infrastructure and data	Data availability
	Government procurement of advanced technology
	Data/AI capability (in government)
Skills and education	Technology skills
	Private sector innovation capability
	Number of AI startups
Government and public services	Digital public services
	Effectiveness of government
	Importance of IT to government's vision of the future

Source: Government Readiness AI Index, 2021.

ITUEvents



AI for Good

Global Summit

*Accelerating the United Nations
Sustainable Development Goals*

30 - 31 May 2024
Geneva, Switzerland

aiforgood.itu.int





Our Goal: Deploy AI to achieve the SDGs

Capitalizing on the immense volume of data available to and use AI to tackle the world's greatest challenges



- **Detect, present** and help **scale-up use cases** for AI enabling the 17 SDGs
- The use of AI for Sustainable Development Goals will allow to:
 - **Monitor** progress towards the achievement of SDG
 - **Simulate** implications
 - **Predict** outcomes of measures taken
 - **Recommendations** for policy makers

AI for Sustainable Development Goals (AI4SDGs) Think Tank

A global collection of AI projects and proposals that impacts UN Sustainable Development Goals, both positively and negatively. The goal is to promote the positive use of AI for Sustainable Development, and to investigate on the negative impact of AI on Sustainable Development. Detailed evaluation on each project is provided based on our rating scheme. You are welcome to share your project to the world and get evaluated by submitting your project or proposal information here.

[BROWSE BY GOALS](#)[SHARE YOUR PROJECT](#)

NDM: Morocco's 2035 Vision

THE NEW DEVELOPMENT MODEL: SHAPING OUR NATION NOW AND FOR THE FUTURE

A PROSPEROUS ECONOMY

Creating wealth and quality jobs for all

REINFORCED SKILLS

Developing skilled and talented citizens that take charge of their life

AN INCLUSIVE MOROCCO

Reinforcing inclusion and social justice

A SUSTAINABLE MOROCCO

Preserving resources in the territories

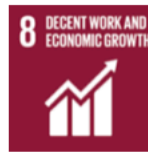


Source: New Development Model, Morocco, 2022.

Morocco's 2035 objectives mapping with UN SDGs

A PROSPEROUS ECONOMY

Creating wealth and quality jobs for all



REINFORCED SKILLS

Developing skilled and talented citizens that take charge of their life



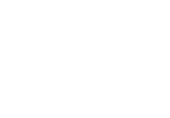
AN INCLUSIVE MOROCCO

Reinforcing inclusion and social justice



A SUSTAINABLE MOROCCO

Preserving resources in the territories



AI Enabled NDM

A PROSPEROUS ECONOMY
Creating wealth and quality jobs for all

AN INCLUSIVE MOROCCO
Reinforcing inclusion and social justice



REINFORCED SKILLS
Developing skilled and talented citizens that take charge of their life

A SUSTAINABLE MOROCCO
Preserving resources in the territories

AI Challenges & Risks



Linkedin, 2024

Six ways to attack an AI system.

Are your AI applications prepared for them?



Poisoning

AI poisoning is a tactic where attackers manipulate the data used to train artificial intelligence (AI) models, causing these models to produce incorrect results or become unreliable. Attackers can introduce subtle errors into training data, such as mislabeling images or biased information, or embed hidden triggers that cause the AI to act unexpectedly when activated. This manipulation can occur intentionally by bad actors, accidentally by use of biased or poor-quality data, or even during normal use if the AI continues to learn from manipulated input or AI content ("feedback loops").



Trojan Horse

With this form of attack, bad actors secretly insert harmful code into AI models, especially large language models, before companies use them, expecting that they cannot check what is hidden inside these models when they obtain them from open sources or buy them. Once these tampered models are used, the hidden malicious code may be activated in one way or another, acting like a trojan horse and using, for instance, unprotected systems (e.g., third-party tools with elevated privileges or insecure browsers) to launch attacks from within a company.



Prompt Injection

Prompt Injection attacks involve tricking an AI system by entering malicious commands instead of normal input. These commands can manipulate the AI to perform unintended actions, like revealing sensitive data or the secret "system prompts" of an AI system, turning off safety controls, or even taking control of other systems that process the output generated by an AI system that is being misused by an attacker. Malicious commands can be included in prompts, but also in documents that a user may upload to an AI system for analysis, resulting in manipulated output.



Sponge Attack

Sponge attacks target AI systems by overwhelming them with complex or large inputs, like a sponge soaking up their computing power. This can slow down or even damage a system. Attackers may do so by crafting inputs that are hard to process, causing the AI to use excessive energy or memory. Such harmful input may be included in a model during the training phase, making the system vulnerable from the start, or they are added later on. This can lead to delays, damage, or safety risks, for example where AI system must remain responsive at all times (e.g., in autonomous vehicles).



Model & Data Theft

Attackers target AI systems to uncover secret data contained in them or how an AI or its model was built. They might trick the AI into revealing if certain data was used in its training or infer private details from the AI's responses. One method does so by testing the system with real data to determine whether it recognizes it with certainty, indicating that it has already seen it during training. Another approach involves flooding the system with specific questions to replicate its logic. These tactics may not only expose sensitive or proprietary information but can lay groundwork for more advanced attacks.



Deception

Attackers can trick AI systems that rely on pattern recognition by using manipulated input to trigger certain (false) responses. For example, if an AI relies on image recognition to classify objects (e.g., speed limit signs), the attacker may use visual elements (e.g., certain stickers on a sign) that may even be invisible to a human to cause the AI to incorrectly assess the object. This may also work with face recognition. In a "white-box" attack the attacker has inside knowledge of the model, whereas in a "black-box" attack, the attacker figures out how to deceive the AI through trial and error.

Reforms, regulations, policies and practices essential to proactively manage AI to ensure responsible, transparent and ethical innovation and avoid misuse and disruptions to human welfare

- Technology is a powerful tool. However, the nature of technology change drives random and “unknowable” changes in production and consumption processes
- AI advancements require the incorporation of ethics and responsibility (both individual and collective) for making policies that harnesses technology in a sustainable manner
- Intelligent machines can make decisions that are more “efficient” but lack of ethics and misuse could cause serious harm to human welfare and requires to be monitored constantly
- Failure to do so results in disruptions in governance of economies and societies and could be seriously catastrophic.
- Therefore AI requires greater engagement and responsible planning from policy makers, industries, researchers and individuals throughout all processes from inception to end use.
- The circular model presents various mechanisms by which technological innovations can enable sustainable development.
- The agents in society – individuals, industry, institutions and community – need to work collaboratively to ensure that technology changes are driven towards the goal of **Enhancing Human (individual and societal) Prosperity, Welfare and Wellbeing**

Global AI Regulations Landscape



**legal
nodes**

March 2024



EU AI Act

Proposal for a

Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

2021/0106 (COD)

European
Commission

PRESIDENT BIDEN SIGNED AN EXECUTIVE ORDER ON AI TO:

SET NEW STANDARDS FOR AI
SAFETY AND SECURITY



PROTECT AMERICANS'
PRIVACY



ADVANCE EQUITY AND
CIVIL RIGHTS



SUPPORT
WORKERS



PROMOTE INNOVATION AND
COMPETITION



ADVANCE AMERICAN
LEADERSHIP ABROAD



ENSURE RESPONSIBLE AND
EFFECTIVE GOVERNMENT
USE OF AI



Concluding Remarks

- Promote education and training in AI to qualify human resources and increase the level of technical awareness in society.
- Support research and innovation by allocating more investments in IA, which contributes to developing innovative solutions that serve various economic sectors.
- Encourage investment in technology companies by creating a suitable investment environment
- Develop the technological infrastructure to enable AI applications in various fields, such as health, education, transportation, etc.
- Strengthen international cooperation in AI through exchange expertise, knowledge and strategic partnerships.

Thank you for your attention

Q & A