OECD.AI
Policy Observatory

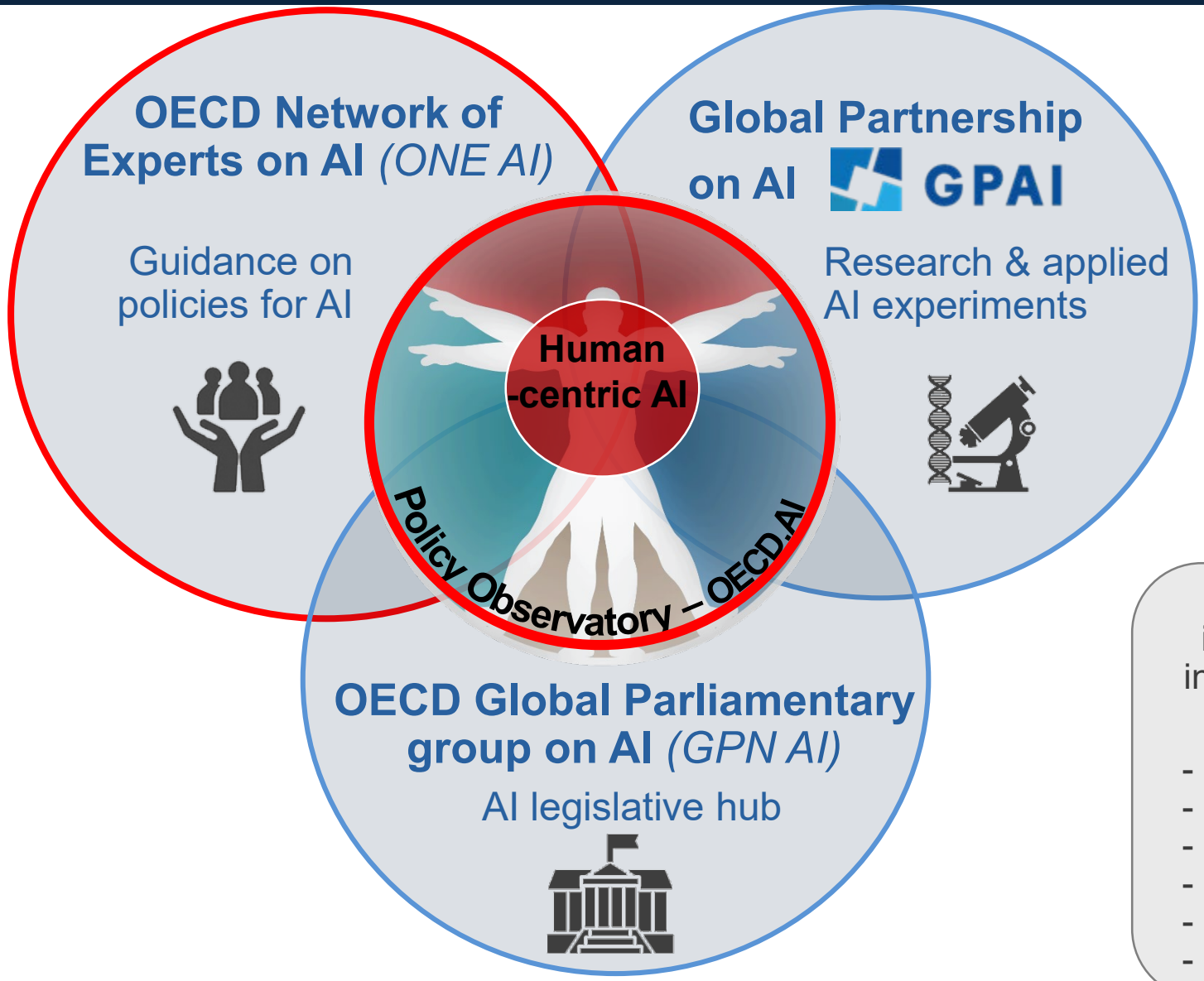
OECD work to help implement trustworthy, human-centric AI

2 December 2020

Andy Wyckoff
OECD Director for Science, Technology and Innovation

OECD

Complementary expert groups – different memberships & mandates

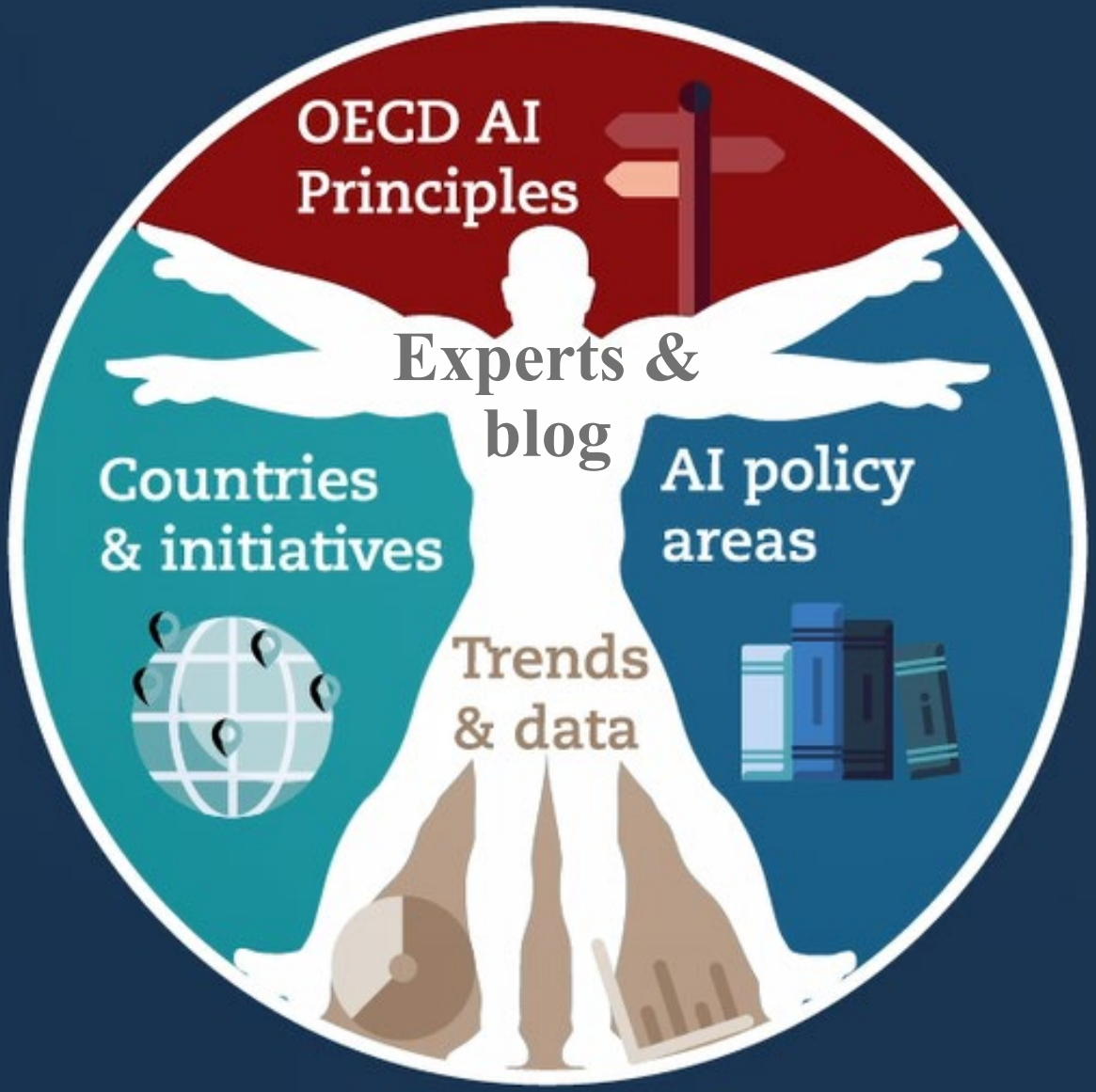


PARTICIPATION in complementary initiatives by partner institutions

- EC
- G20
- UNESCO, UN
- CoE
- IDB, CAF
- ISO, IEEE…

OECD.AI

Shape and share public policies for responsible, trustworthy and beneficial AI



OECD Policy C

OECD.AI Policy Observatory

OECD.org Going Digital Toolkit

Experts & blog | AI Principles | Policy areas | Trends & data | Countries & initiatives | About

OECD AI Policy Observatory

We provide data and multi-disciplinary analysis on artificial intelligence. Our diverse global community of partners makes this platform a unique source of information and dialogue on AI.

Shape and share public policies for responsible, trustworthy and beneficial AI

OECD AI Principles

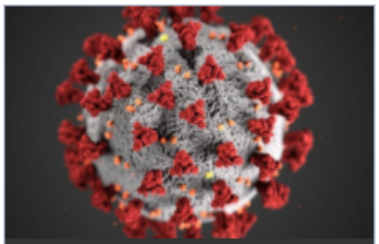
Wondering what standards to apply to AI policies and practices?

The AI Principles give guidance for a human-centred trustworthy approach.

AI Policy areas

Explore how AI affects everything from transport to jobs and education.

Find out about AI's impact on work innovation, productivity and skills.



COVID 19

AI-powered live news, data-viz, data for AI

Countries & initatives

Explore over 300 AI policy initiatives from over 60 countries.

Find the latest AI initiatives from business, technologists and others.

Trends & data

Keep up with the latest AI developments and trends.

Explore live news, data and research from the OECD and its partners.

Video

The Launch of the OECD AI Policy Observatory

Latest posts on the OECD AI Wonk

Technical community

COVID-19 and beyond: Elements of certainty can make AI ecosystems trustworthy

Standardization, certification and appropriate governance structures can help to establish trust for AIS.

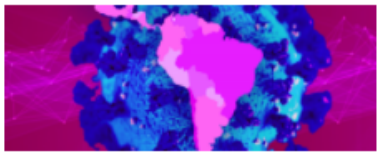
November 16, 2020 — 8 min read

Technical community

How private applied AI R&D labs can make AI work for mid-size businesses

Private applied AI R&D labs are the key to keeping European mid-size businesses (MSBs) competitive.

October 23, 2020 — 8 min read



Intergovernmental

Webinar: How can Artificial Intelligence help Latin America and the Caribbean fight the COVID-19 virus?

A webinar on the latest developments on how artificial intelligence has helped to combat COVID-19.

October 20, 2020 — 3 min read

OECD.AI Policy Observatory

OECD.org | Going Digital Toolkit

Experts & blog | AI Principles | Policy areas | Trends & data | Countries & initiatives | About

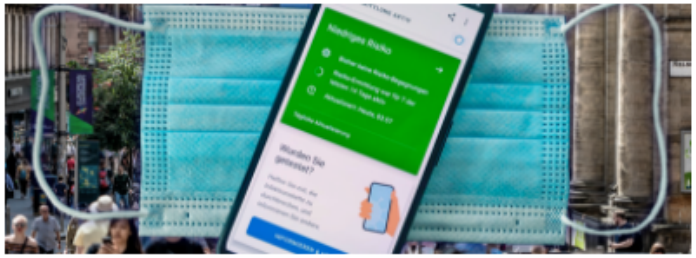
Home > The AI Wonk

The AI Wonk

See oecd.ai/wonk

2 new expert contributions per week

Not all intelligence is artificial. Keep yours real with the AI Wonk blog.



Technical community

COVID-19 and beyond: Elements of certainty can make AI ecosystems trustworthy

Standardization, certification and appropriate governance structures can help to establish trust for AIS.

November 16, 2020 — 8 min read

Intergovernmental

Webinar: How can Artificial Intelligence help Latin America and the Caribbean fight the COVID-19 virus?

October 20, 2020 — 3 min read

Technical community

How do we govern the complex intersection between AI and cybersecurity?

October 1, 2020 — 9 min read

Technical community

How private applied AI R&D labs can make AI work for mid-size businesses

October 23, 2020 — 8 min read

About the AI Wonk

The AI Wonk is a space where the OECD Network of Experts on AI (ONE AI) and guest contributors share their experiences and research. It is an ongoing conversation about the OECD AI Principles and how to best share and shape trustworthy AI policies that benefit individuals, communities and economies.

Contributors

All contributors

Benoît Bergeret
indust.ai
Founding CEO · ONE AI working group on Policies for AI
See all posts

Dr. Clara Neppel
IEEE European Business Operations
Senior Director · ONE AI working group on the Classification of AI
See all posts

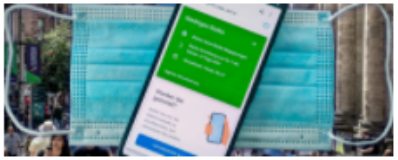
Nicolas Miailhe
The Future Society
Co-Founder · ONE AI working group on Implementing Trustworthy AI
See all posts

All stakeholder types

Latest posts

Civil society

iHuman: a virtual screening of the film



Technical community

COVID-19 and beyond: Elements of

Intergovernmental

An open call for input from GPAI's Data

Transparency and explainability (Principle 1.3)

This principle is about transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes.

Rationale for this principle

The term transparency carries multiple meanings. In the context of this Principle, the focus is first on disclosing when AI is being used (in a prediction, recommendation or decision, or that the user is interacting directly with an AI-powered agent, such as a chatbot). Disclosure should be made with proportion to the importance of the interaction. The growing ubiquity of AI applications may influence the desirability, effectiveness or feasibility of disclosure in some cases.

Transparency further means enabling people to understand how an AI system is developed, trained, operates, and deployed in the relevant application domain, so that consumers, for example, can make more informed choices. Transparency also refers to the ability to provide meaningful information and clarity about what information is provided and why. Thus transparency does not in general extend to the disclosure of the source or other proprietary code or sharing of proprietary datasets, all of which may be too technically complex to be feasible or useful to understanding an outcome. Source code and datasets may also be subject to intellectual property, including trade secrets.

An additional aspect of transparency concerns facilitating public, multi-stakeholder discourse and the establishment of dedicated entities, as necessary, to foster general awareness and understanding of AI systems and increase acceptance and trust.

Explainability means enabling people affected by the outcome of an AI system to understand how it was arrived at. This entails providing easy-to-understand information to people affected by an AI system's outcome that can enable those adversely affected to challenge the outcome, notably – to the extent practicable – the factors and logic that led to an outcome. Notwithstanding, explainability can be achieved in different ways depending on the context (such as, the significance of the outcomes). For example, for some types of AI systems, requiring explainability may negatively affect the accuracy and performance of the system (as it may require reducing the solution variables to a set small enough that humans can understand, which could be suboptimal in complex, high-dimensional problems), or privacy and security. It may also increase complexity and costs, potentially putting AI actors that are SMEs at a disproportionate disadvantage.

Therefore, when AI actors provide an explanation of an outcome, they may consider providing – in clear and simple terms, and as appropriate to the context – the main factors in a decision, the determinant factors, the data, logic or algorithm behind the specific outcome, or explaining why similar-looking circumstances generated a different outcome. This should be done in a way that allows individuals to understand and challenge the outcome while respecting personal data protection obligations, if relevant.

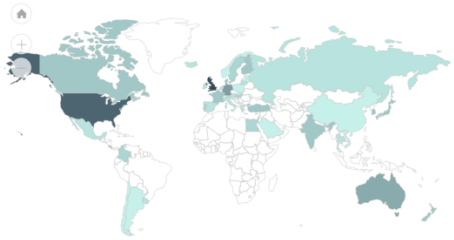
Related online news from EventRegistry

News language: English only

- Coronavirus: NSE gives more time to brokers for submission of reports — Apr 9, 2020
- Compliance: Your New Best Friend? — Jan 27, 2020
- Ajman Department of Ports and Customs launches artificial and business intelligence (AI/BI) powered services — Ajman — Jan 14, 2020
- Here's how DBS is trying to fend off Financial crimes — Nov 15, 2019

Related AI policy initiatives

Choose visualization: By initiative count



European Union (13)

Related OECD publications

Choose visualization: From newest to oldest

7 results

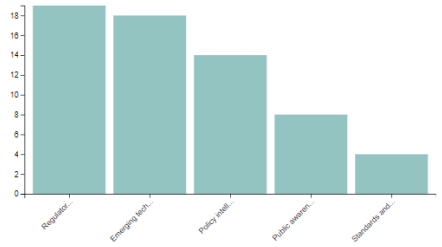
- 2019 - Artificial Intelligence in Society: Public policy considerations
- 2019 - Artificial Intelligence in Society: The technical landscape
- 2019 - Hello, World: Artificial intelligence and its use in the public sector
- 2019 - Scoping the OECD AI principles : Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)

Related recent scientific research

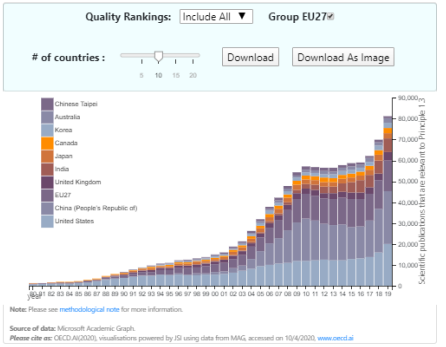
Research status: Published

- DeepXplore: automated whitebox testing of deep learning systems — Oct 24, 2019
- XLNet: Generalized Autoregressive Pretraining for Language Understanding — Dec 8, 2019
- Deep Supervised Hashing for Fast Image Retrieval — Sep 1, 2019
- One Pixel Attack for Fooling Deep Neural Networks — Oct 1, 2019

Related types of policy instruments used



Related scientific research by country



AI & employment

AI is widely expected to change the nature of work as it diffuses across sectors. It will complement humans in some tasks, replace them in others and also generate new types of work. This section showcases the latest on AI and employment.

Related OECD publications

Choose visualization: From newest to oldest

12 results

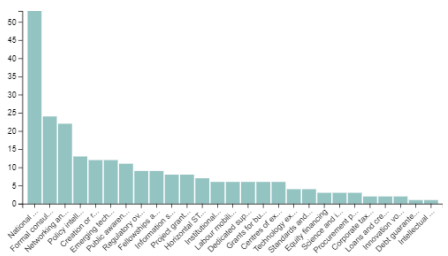
- 2019 - Measuring the Digital Transformation: Adaptability
- 2019 - OECD Employment Outlook 2019 : The Future of Work
- 2019 - OECD Skills Outlook 2019: Thriving in a Digital World
- 2019 - Statistical profiling in public employment services : An international comparison

Related online news from EventRegistry

News language: English only

- Don't fear a 'robot apocalypse' - tomorrow's digital jobs will be more satisfying and higher-paid
  Feb 27, 2020
- Hubtek offers robotic process automation
  Feb 25, 2020
- Automation, climate change, AI: schools prepping students for jobs of the future
  Alberta Jan 2, 2020
- Reskilling the UK in the face of AI growth
  Dec 9, 2019

Related types of policy instruments used



Related top concepts in research

Download | Download As Image

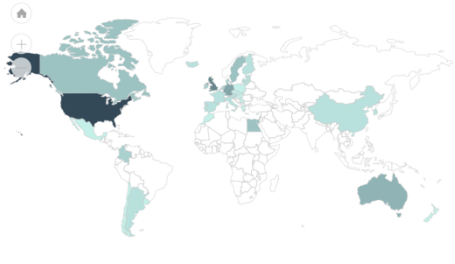
Note: this chart takes into account research in employment and AI from 1980 onwards. Please see methodological note for more information.

Source of data: Microsoft Academic Graph.

Please cite as: OECD.AI(2020), visualisations powered by JSI using data from MAG, accessed on 10/4/2020, www.oecd.ai

Related AI policy initiatives

Choose visualization: By initiative count



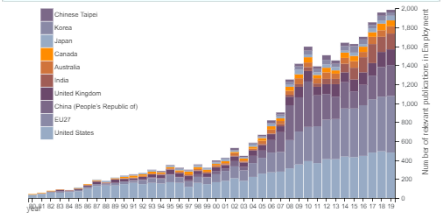
European Union (3)

Related scientific research by country

Quality Rankings: Include All | Group EU27

# of countries: 5 10 15 20

Download | Download As Image



Note: Please see methodological note for more information.

Source of data: Microsoft Academic Graph.

Please cite as: OECD.AI(2020), visualisations powered by JSI using data from MAG, accessed on 10/4/2020, www.oecd.ai

Related recent scientific research

Research status: Published

- A scalable pipeline for designing reconfigurable organisms
  Jan 13, 2020
- Gender Differences in Earnings of Early- and Midcareer Pediatricians
  Oct 1, 2019
- Genetic design automation for autonomous formation of multicellular shapes from a single cell progenitor
  Oct 16, 2019
- Visual Interaction with Deep Learning Models through Collaborative Semantic Inference
  Jan 1, 2020

Programme on Work, Innovation, Productivity and Skills (AI-WIPS)

The OECD.AI AI-WIPS programme is supported by Germany (with KI-Observatorium).

It analyses the impact of AI on **labour market, skills and social policy.**

5 central themes:

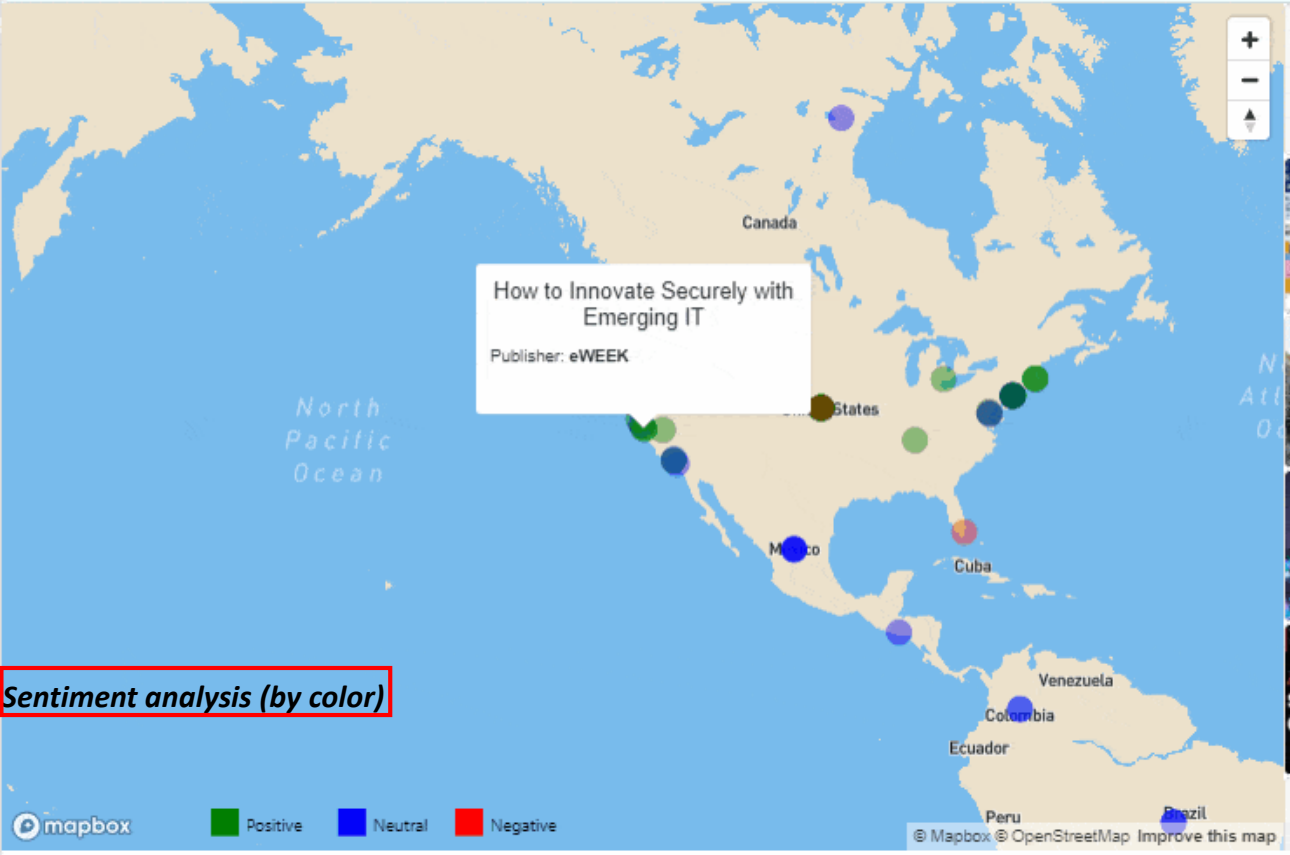
- *Identifying and classifying AI Systems* (including through the ONE AI WG on AI classification)
- *Assessing AI and robotics capabilities*
- *AI and workforce skills*
- *Assessing the speed of AI diffusion*
- *The impact of AI on labour markets and society*

Live data from partners

This section leverages live data from partners to show timely trends about where, how and at what rate AI is being developed and used, and in which sectors.

AI news | AI research | AI jobs and skills

Country: All countries Speed: slow normal fast NewsType: articles Reading mode Show English only Paused



Sentiment analysis (by color)

eWEEK
How to Innovate Securely with Emerging IT
Fri Oct 02 2020, 14:50

SankeiBiz サンケイビズ
大阪大学が新型コロナウイルス感染症等に関するSNSによるこころの相談の充...
Fri Oct 02 2020, 14:53

Računalniške novice
Spoznajte edini mesti na svetu, ki svoje algoritme odpirata ljudem
Fri Oct 02 2020, 15:07

MIT Technology Review
A VR film/game with AI characters can be different every time you...
Fri Oct 02 2020, 15:10

@palestinechron
Why is the World Going to Hell? Netflix's The Social Dilemma Tells...
Fri Oct 02 2020, 15:12

TechRepublic
MIT Report: How AI changes the

Live data from partners

This section leverages live data from partners to show timely trends about where, how and at what rate AI is being developed and used, and in which sectors.

AI news | AI research | AI jobs and skills | AI search trends | COVID-19 research

Year: 1980 1990 2000 2010 2020 Resume

Publications Per Capita ☐ Cumulative ☐ Estimated 2020 value ☑

Y-axis: GDP Per Capita

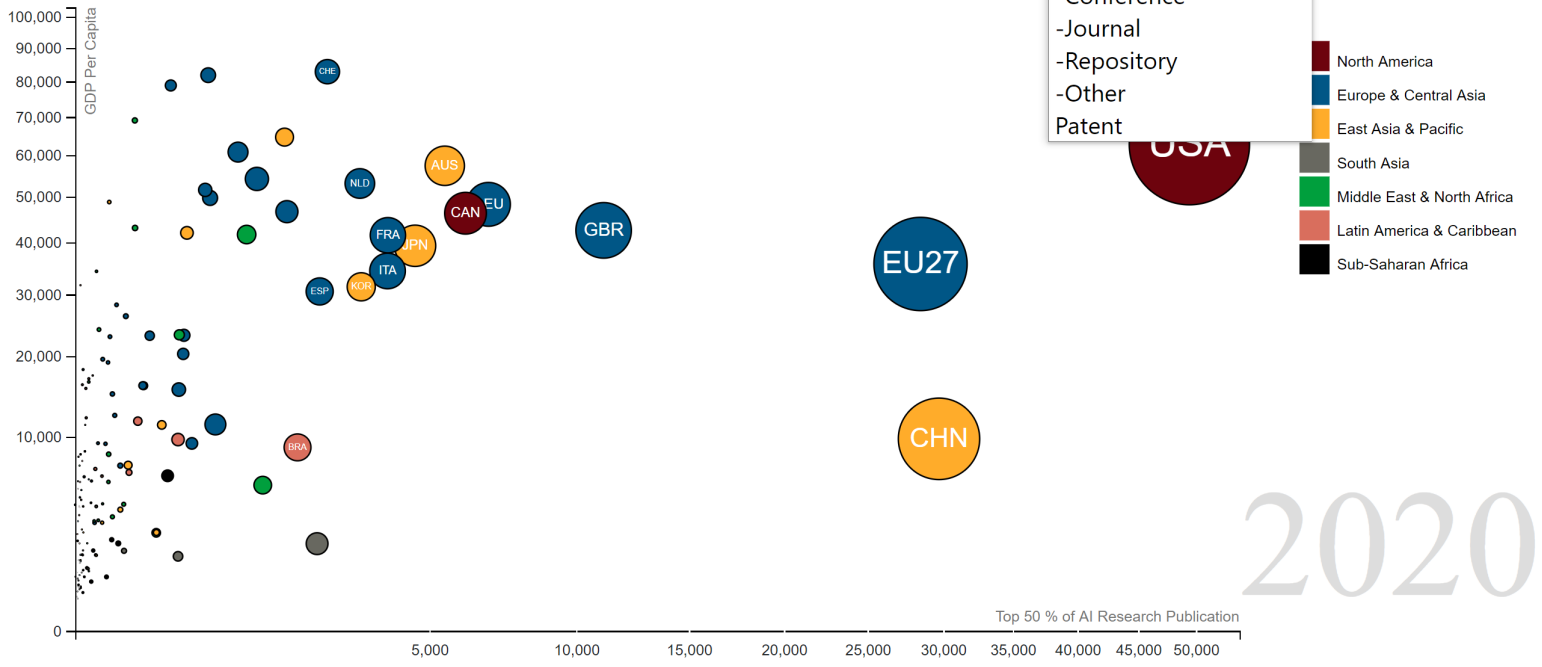
Quality Rankings: Top 50%

Publication Type: Research Publications

- Research Publications
- -Book
- -BookChapter
- -Conference
- -Journal
- -Repository
- -Other
- Patent

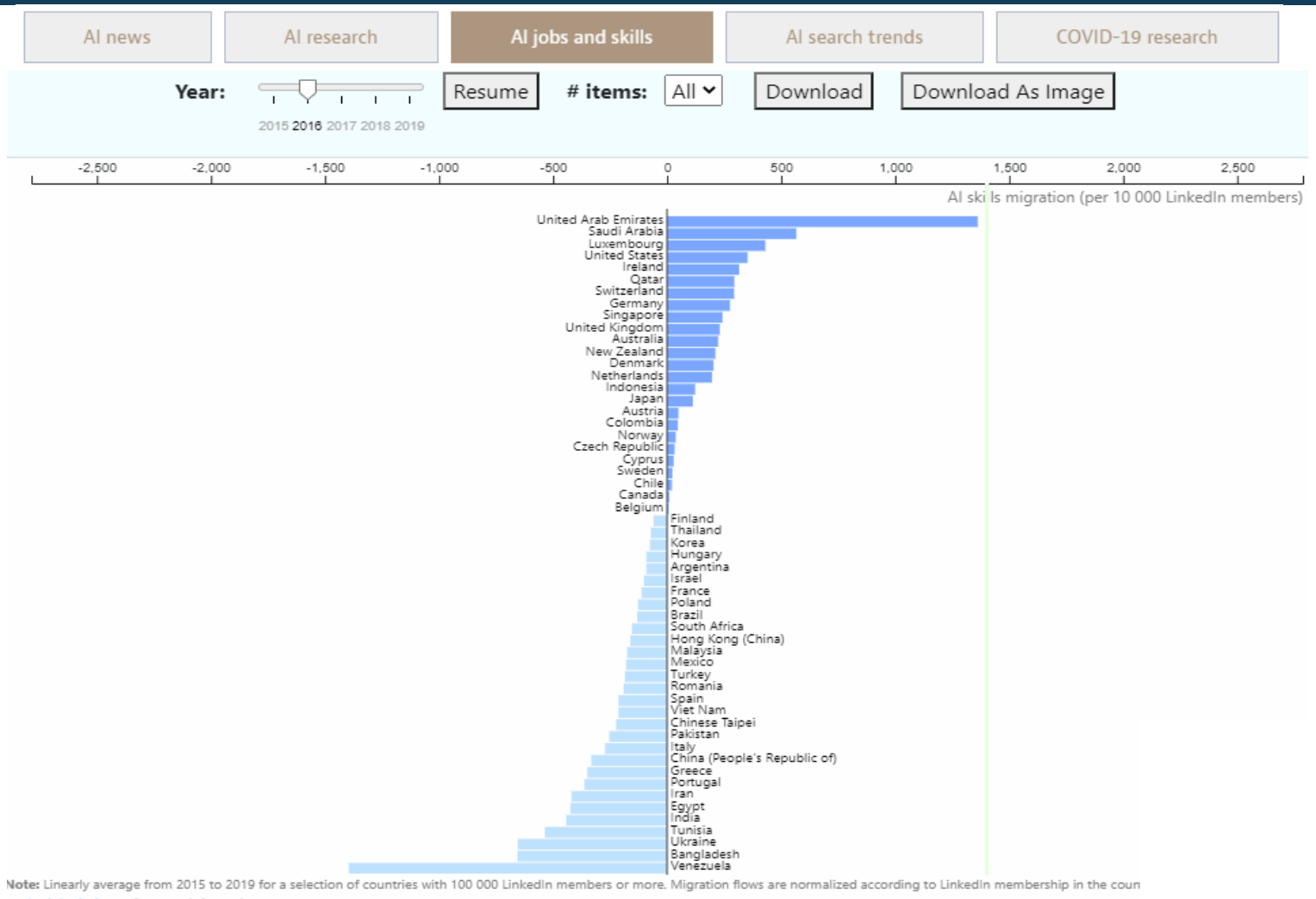
Download

Download As Image



Live data from partners

This section leverages live data from partners to show timely trends about where, how and at what rate AI is being developed and used, and in which sectors.



Note: Linearly average from 2015 to 2019 for a selection of countries with 100 000 LinkedIn members or more. Migration flows are normalized according to LinkedIn membership in the coun

Live data from partners

This section leverages live data from partners to show timely trends about where, how and at what rate AI is being developed and used, and in which sectors.

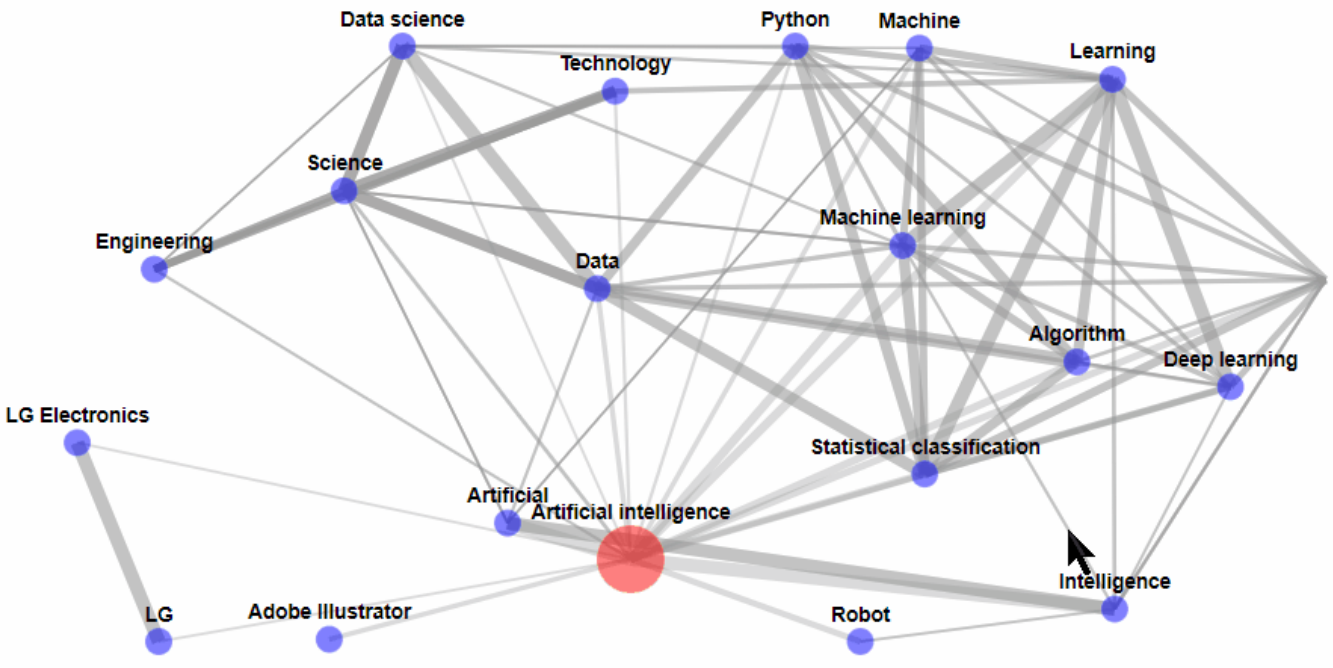
AI news | AI research | AI jobs and skills | **AI search trends** | COVID-19 research

Country: All countries

# of terms: 20 40 60 80 100

Date: 2015-01 2015-11 2016-09 2017-07 2018-05 2019-03 2020-08

Weighted percentile: 0 20 40 60 80 100



Note: Please see methodological note for more information.

OECD.AI Policy Observatory

Work in progress

- AI skills
- AI software development
- AI venture capital
- AI research by gender
- AI education
- Computing power
- Mentions of AI in financial statements

OECD.AI Policy Observatory

OECD.org | Going Digital Toolkit

Experts & blog | AI Principles | Policy areas | Trends & data | Countries & initiatives | About

Home > Countries & Initiatives > National strategies & policies

National AI policies & strategies

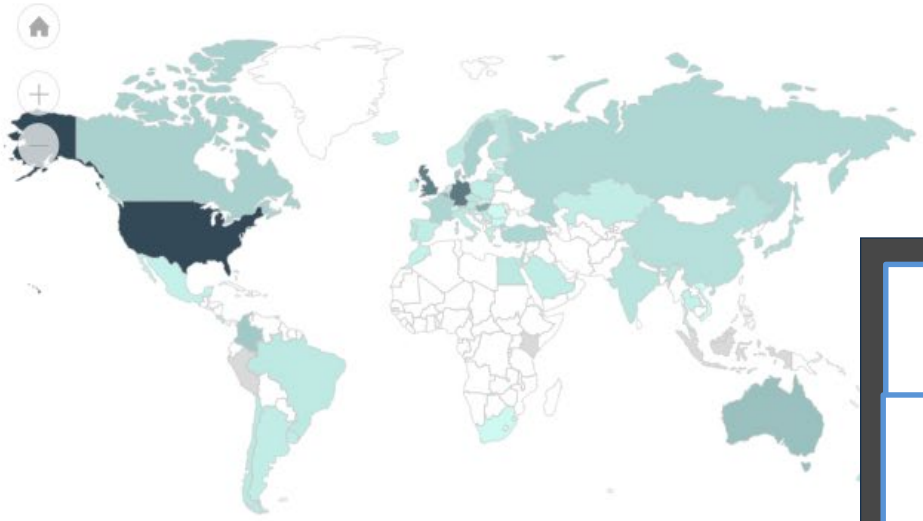
This section provides a live repository of over 300 AI policy initiatives from 60 cou... ...y instrument or a group targeted by the policy.

Please send us AI legislative updates at ai@oecd.org

Countries & territories | Policy instruments | Target Groups

European Commission

Choose visualization: By initiative count



European Union (28)

60

Countries, territories and the EU

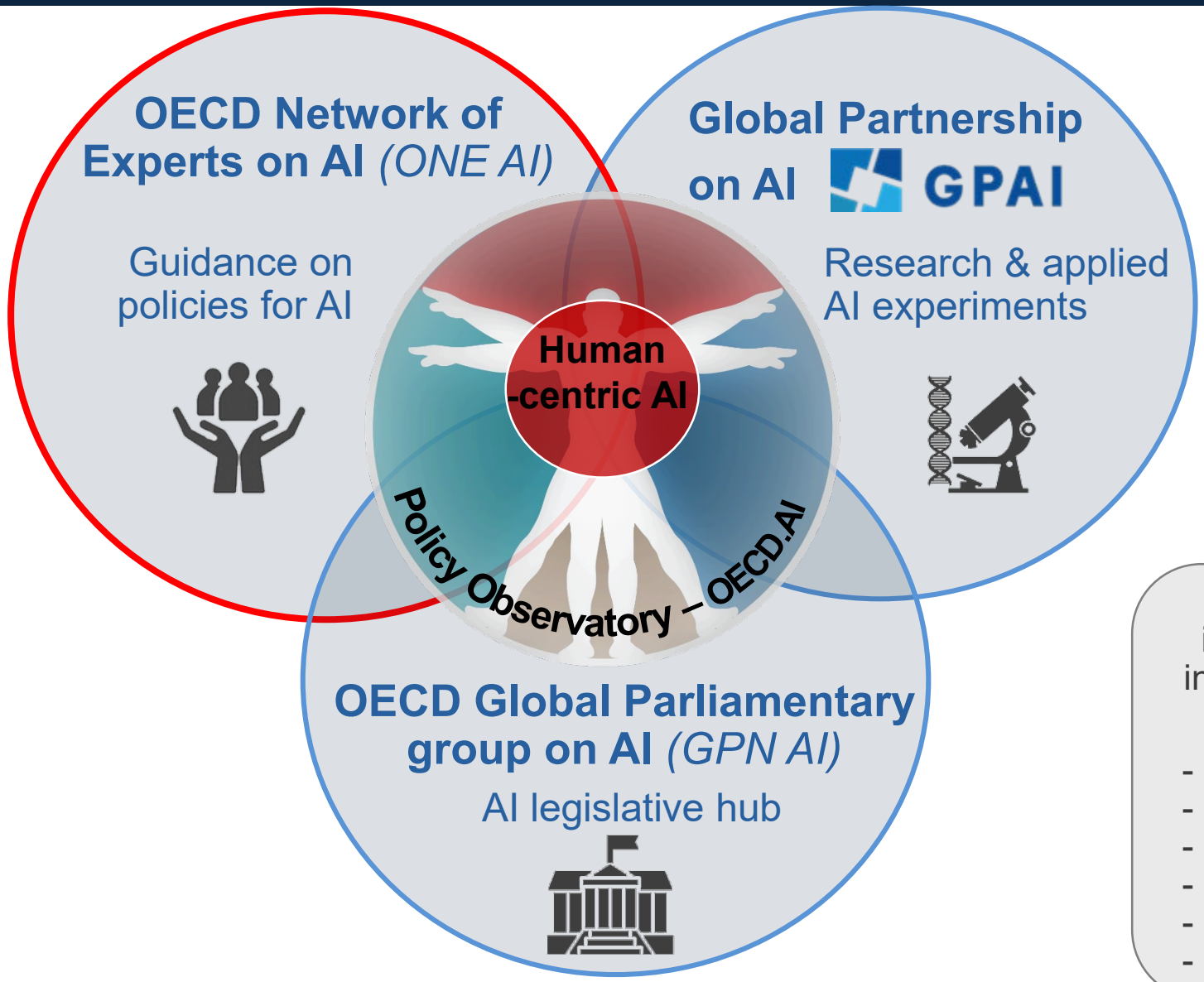
600

national AI policies, initiatives, instruments in

new Q1 2021 version

New focus on emerging AI regulations across these 60 countries: both soft and hard law s

Complementary expert groups – different memberships & mandates



PARTICIPATION in complementary initiatives by partner institutions

- EC
- G20
- UNESCO, UN
- CoE
- IDB, CAF
- ISO, IEEE…

Implementing the AI Principles in POLICIES – ONE AI priorities

OECD.AI Policy Observatory

What types of AI systems raise what types of policy issues?

Classifying AI systems

How to measure national AI compute?

10 Principles for Trustworthy AI

Values-based principles

Socio-economic & environmental impacts, Human-centred values and fairness
Transparency, explainability
Robustness, security, safety
Accountability

What tools help implement trustworthy AI

Tools for Trustworthy AI

National Policies

Investing in research
Data, algorithms, **AI Compute**
Enabling policy environment
Jobs, skills, transitions
International cooperation

National AI policies

What have we learned so far about (national) AI policies?

OECD.AI Policy Observatory

ONE AI focus areas

Working Group / Task Force	Goals and deliverables
AI classification	user-friendly framework to classify and help policy makers assess the different policy considerations associated with different types of AI systems.
Tools for trustworthy AI	practical guidance on tools that can help AI actors and decision-makers implement trustworthy AI.
AI policies	practical guidance on designing, implementing & monitoring national AI policies.
AI compute NEW	framework for understanding and measuring key components of domestic AI computing capacity.

OECD.AI Policy Observatory

Global Partnership on AI (GPAI)

Launched June 2020, 15 founders, aligned on OECD AI principles, more likeminded members to join

ABOUT GPAI

GPAI brings together leading experts from industry, civil society, governments, and academia to bridge the gap between theory and practice on AI -- through cutting-edge research and applied activities.

GPAI EXPERT WORKING GROUPS

1) Responsible AI
   -subgroup on AI & Pandemics
2) Data Governance
3) Future of Work
4) Innovation and Commercialisation

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

OECD.AI
Policy Observatory

Thank you

Email: **ai@oecd.org**

Follow: **OECD.AI LinkedIn page**
http://www.linkedin.com/company/oecd-ai/

Sign up to the **newsletter**

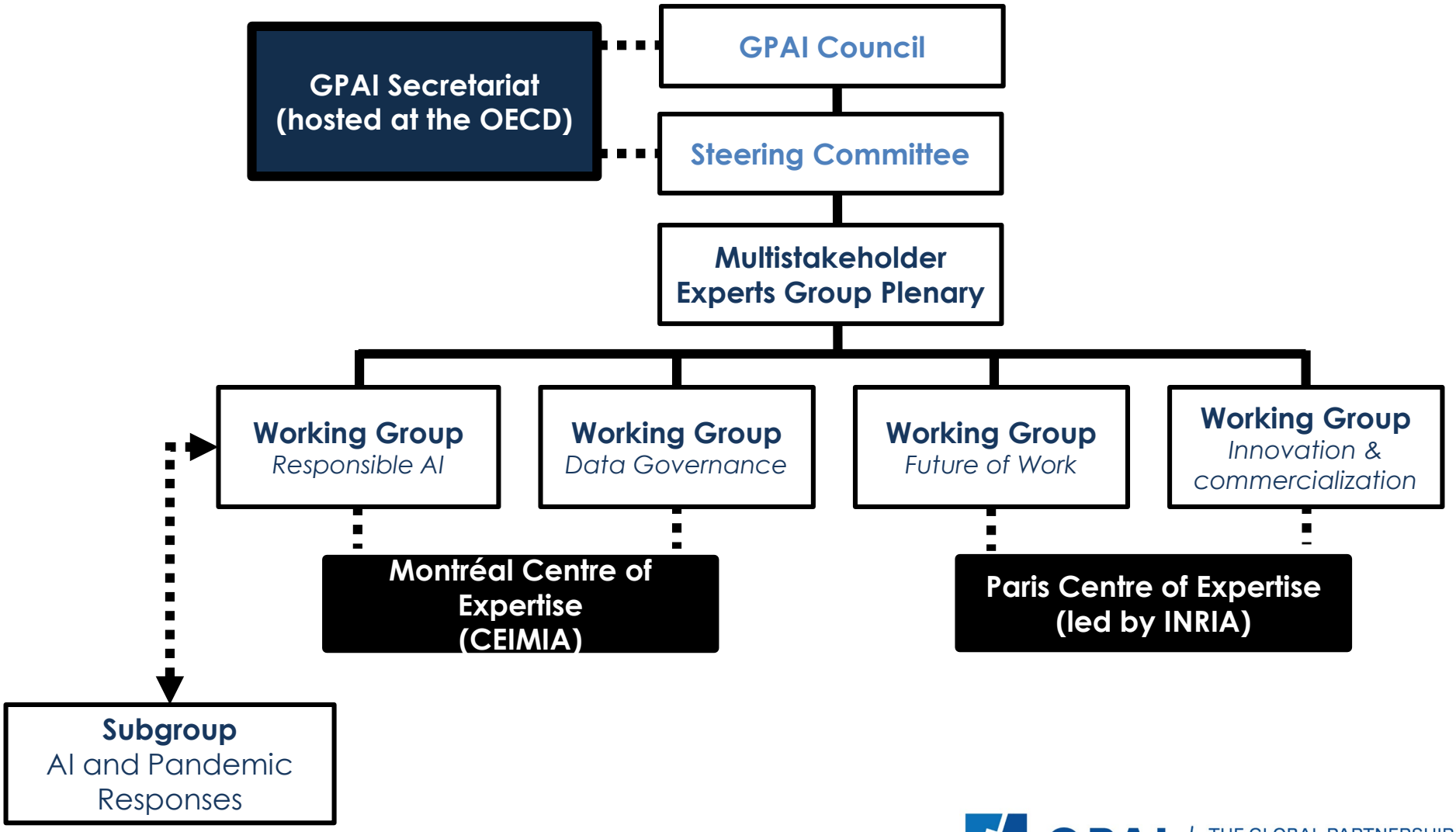
OECD

OECD.AI
Policy Observatory

BACK-UP

OECD.AI Policy Observatory

Global Partnership on AI (GPAI)



GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

OECD AI Systems Classification Framework

Objective:

- provide a structure to assess and classify AI systems according to their impact on public policy in areas covered by the OECD AI Principles.

Key points:

- The framework is simplified and user-friendly rather than exhaustive.
- The robustness and applicability of the present framework will be tested in late 2020 / early 2021 and adjustments made if needed.
- The 10 OECD AI Principles are used to structure the analysis of policy considerations associated with each dimension and sub-dimension

Values-based principles for all AI actors

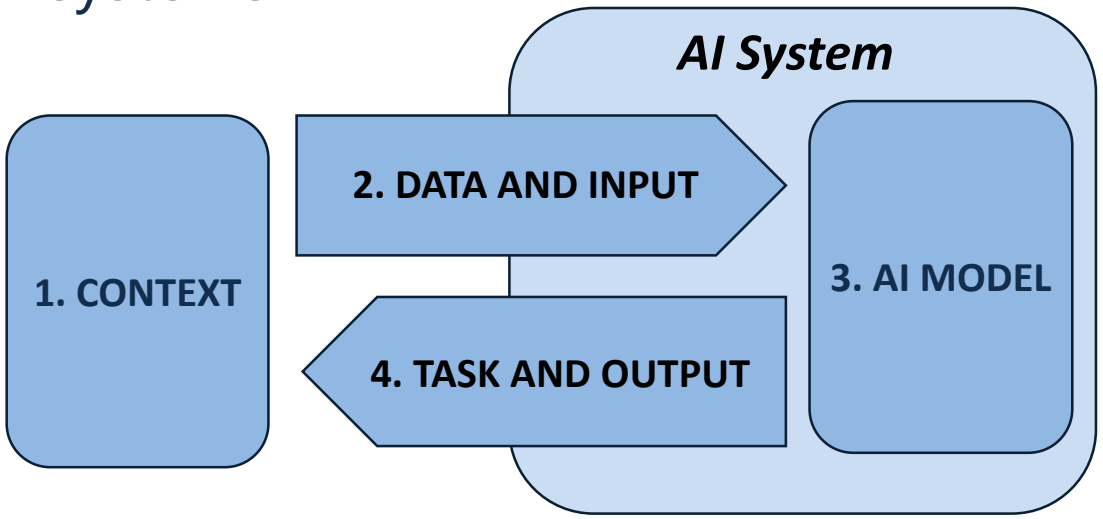
1.1. Economic, social and environmental impact
1.2. Human rights including privacy, fairness
1.3. Transparency, explainability
1.4. Robustness, security, safety
1.5. Accountability

Recommendations to policy makers

2.1. Investment in research
2.2. Data, compute, technologies
2.3. Enabling policy and regulatory environment
2.4. Jobs, automation, skills
2.5. International cooperation

Overview of AI Systems Classification Framework

A user-friendly framework to navigate policy implications of different *types* of AI systems



4 key dimensions:

1. **Context**, including sector (healthcare, etc.), impact and scale
2. **Data and input**, including data collection, personal nature of data
3. **AI model (technologies)**, incl. model type and model building process
4. **Task and output**, incl. AI system's task (e.g., recognition, personalisation, etc.) and action autonomy

OECD.AI Policy Observatory

Tools to implement trustworthy AI

Type of approach	Type of tool
Process-related approaches	Guidelines
	Governance frameworks
	Product development / lifecycle tools
	Risk management
	Sector-specific codes of conduct
	Process standards
Technical approaches	Technical validation approaches
	Technical documentation
	Technical standards
	Toolkits / toolboxes / software tools
Capacity / awareness building approaches	Educational material
	Change management processes

OECD.AI Policy Observatory

Tools to implement trustworthy AI *(WIP)*

Tool properties [*what the tool is*]

- Type of approach
- Type of tool
- Tool maturity
- Tool design
- Tool usage
- Tool application

Tool functionality [*what the tool does*]

- Relevant to which AI Principle
- Relevant to which AI lifecycle stage
- Helps to assess impact
- Helps to evaluate implementation
- Includes accountability mechanisms
- Encourages consultations with stakeholders
- Includes sharing mechanisms

Implementation incentives [*why use the tool*]

- Incentives to use
- Enforcement mechanisms

OECD.AI Policy Observatory

Develop a tool classification framework *(WIP)*

Tool properties [*what the tool is*]	Type of approach		e.g. see Table A A.1
	Type of tool		e.g. see Table A A.1
	Tool maturity		e.g. project stage; in development; running code
	Tool design	Tool developer	e.g. academia; business; civil society; government; intergovernmental; multi-stakeholder; NGO; technical community; trade union
		Tool maintenance	e.g. tool is maintained / sustainable
	Tool usage	Target audience of the tool	e.g. designers; developers; workers; etc.
		Resource requirements	e.g. infrastructure; data; skill; budget requirements...feasibility for SMEs, developing countries, ease of implementation, etc.
		Openness of the tool	e.g. closed; open; other
	Tool application	Sector specificity	e.g. sector specific; horizontal (applicable to several sectors)
		Neutrality	e.g. technology / platform neutral
		AI specificity	e.g. AI-specific; broader digital issue
		Scalability	e.g. scalable / can be mainstreamed
		Scope	e.g. general; practical
Tool functionality [*what the tool does*]	Relevant to which AI Principle(s)		e.g. Principles 1.1-2.5
	Relevant to which AI system lifecycle stage		e.g. planning & design; data collection & processing; model building & interpretation; verification & validation; deployment; operation & monitoring
	Helps to assess impact of implementing Principles		e.g. impact; trade-off; cost-benefit analysis
	Helps to evaluate implementation of Principles		e.g. testing; evaluation; monitoring
	Includes accountability mechanisms		e.g. ombudsman; ethical board; compliance; redress mechanisms; clear roles & responsibilities -- responsible actor for system malfunction
	Encourages consultations with affected parties		e.g. on the design and/or implementation of AI systems
	Includes sharing mechanisms / platforms		e.g. to share good practices, lessons learned, code improvements, etc.
Implementation incentives [*why use the tool*]	Incentives to use		e.g. financial; social / reputational; ethical; job quality; regulatory incentives; difficulty of free-riding
	Enforcement mechanisms		e.g. certification; regulation; oversight body

National AI policy cycle: 4 sections

1. Policy design – governance, processes to formulate national AI policies

National AI governance

Horizontal coordination

Consultations

2. Policy implementation

Investing in AI R&D;

Data, compute, software & knowledge;

Regulation, testbeds, documentation;

Automation, skills;

Tools for trustworthy AI:

- *Codes of conduct, guidelines*
- *Standards, software tools, research*
- *Capacity building*

3. Policy Intelligence

Translating AI policies into action plans & targets.

Evaluating implementation of AI policies

Benchmarks and indicators

4. International multi-stakeholder co-operation

IGOs e.g. OECD, EC, IDB, CoE, UNESCO, UN, World Bank.

Cooperation on standards e.g. ISO

Multi-stakeholder initiatives

OECD.AI Policy Observatory

Introducing the ONE AI Compute Task Force

- **National policy imperative to invest in AI**
- **AI is not IT - requires specialized compute stack (HW+SW)**
- **AI compute has so far received less attention**
- **Growing AI compute divide = economic inequity**
- **Impact beyond academia – industry, jobs, public health, etc.**
- **Policy makers need non-technical framework to understand:**
  - What do we have?
  - How does it compare?
  - Is it enough?

AI enablers