

Annex B. Indicators' description

Table A B.1. Description of selected indicators

Indicator description	Context and source	Scoring methodology
Number of "high-quality" AI research publications	<p>Two sources were used for research publication data, Elsevier and OpenAlex. Scopus is Elsevier's expertly curated abstract and citation database, with over 75 million indexed records. The data used to construct OECD.AI visualisations includes scholarly articles, conference proceedings, reviews, book chapters and books from a subset of AI-related resources belonging to Elsevier. More than 600,000 AI scholarly publications are extracted from its archives using core AI keywords such as back-propagation neural network, genetics-based machine learning, cohen-grossberg neural networks, back-propagation algorithm, and neural networks learning.</p> <p>The OpenAlex dataset is a comprehensive, open-source bibliographic database offering extensive information on scholarly works. It includes over 245 million research publications, including journals, conferences, and workshop papers. The data records are tagged with a set of 65 000 topics from Wikidata, covering a range of different subjects.</p>	<p>Scientific publications are ranked based on the Field-weighted Citation Impact (FWCI), which is the ratio of the total citations actually received by a scientific publication and the total citations that would be expected based on the average of the subject field or scientific discipline. A FWCI of 1 means that the publication is cited as much as the average publication in that subject field or scientific discipline. A FWCI of less (more) than 1 means that the publication is cited less (more) than the average publication in that subject field or scientific discipline. In this manner, the FWCI takes into account the differences in research behaviour across disciplines.</p> <p>OECD.AI defines three categories of scientific publications based on their FCWI score:</p> <ul style="list-style-type: none"> • Low impact: $0 < \text{FWCI} \leq 0.5$ • Medium impact: $0.5 < \text{FWCI} \leq 1.5$ • High impact: $\text{FWCI} > 1.5$
Number of AI patents applications submitted	<p>Data refer to AI-related patent applications filed under the PCT, by earliest filing date and location of applicants, using fractional counts. AI-related patents are identified on the basis of their Cooperative Patent Classification (CPC) group and keyword search in patent's title and abstract (OECD, forthcoming).</p> <p>Data is compiled by the OECD.</p>	<p>"AI patents" are patents that:</p> <ul style="list-style-type: none"> - Belong to a "core AI" group: Patents classified under Cooperative Patent Classification (CPC) groups: G06N3 (computing arrangements based on biological models), G06N5 (id. Using knowledge-based models), G06N7 (id. Based on specific mathematical models), G06N20 (machine learning), G06F18 (pattern recognition); and G06T9 (image coding); or - Belong to an "AI-related" group and display an AI keyword: Patents from a selected list of AI-related CPC groups whose title or abstract contain at least one relevant AI keyword. This group is made of technologies applying AI to particular uses (e.g., autonomous driving). Combining keywords and CPC groups helps to prevent overestimation of AI patents by filtering out patents pertaining to those

		<p>technologies but which are not AI per se.</p> <p>Patents which fulfill neither of these conditions are not considered as AI.</p> <p>The OECD Secretariat uses patent data with a two-year lag to ensure the reliability of the indicator, as patent filings are a lagging indicator. Patents can take several years to move from application to publication, which makes drawing conclusions from recent data problematic.</p>
Number of AI models developed	A comprehensive data collection was carried out by screening the websites of cloud providers to track the characteristics of AI models on offer, including different modalities. This data was collected as an effort to measure competition in AI development (André et al., 2025 ^[26]).	Counts were determined by unique combinations of model characteristics across 7 segments: foundation model, variant, version, update date, number of parameters, and context window size.
Number of large-scale models	Models in the Epoch AI dataset have been collected from various sources, including literature reviews, Papers With Code, historical accounts, highly-cited publications, proceedings of top conferences, and suggestions from individuals. The list of models is non-exhaustive, but aims to cover most models that were state-of-the-art when released, have over 1000 citations, one million monthly active users, or an equivalent level of historical significance.	The number of models were counted by country associated with the developing organisation(s). Countries were identified for 97% of models.
GPU clusters	Epoch AI's database of over 500 GPU clusters and supercomputers tracks large hardware facilities, including those used for AI training and inference. The dataset covers an estimated 10–20% of existing global aggregate GPU cluster performance as of March 2025.	GPU clusters were aggregated by the number of H100 equivalents, and filtered for confirmed clusters that had their first operational date on or before the last day of the target year.
Fibre connections per 100 inhabitants	The OECD Broadband Statistics includes broadband-related statistics based on questionnaires received by the OECD from relevant national authorities of OECD members and accession countries.	Fibre connections per 100 inhabitants.
Fixed subscriptions above 100mbps over total advertised fixed broadband plans		Fixed broadband subscriptions exceeding certain thresholds (above 100 Mbps) over total advertised fixed broadband plans. Note that these speed thresholds are the speeds advertised by service providers and not the actual speeds experienced by users.
Datasets available by language	<p>Hugging Face offers a leading open-source library for machine learning (ML) tasks, providing state-of-the-art pre-trained models, datasets, and metrics to support ML application development. It serves as a central repository for sharing and versioning these resources, and users can access and download them for a wide range of ML applications such as NLP, computer vision and reinforcement learning.</p> <p>The World Factbook provides basic information on the history, people, government, economy, energy, geography, environment, communications, transportation, military, terrorism, and transnational issues for 258 world entities.</p>	<p>The number of multilingual and monolingual open datasets are acquired from the Hugging Face public API. The number of datasets containing a language were counted for each language appearing in the Hugging Face API. For a calculation for a given year in the Index, only datasets created at or before that year were included.</p> <p>Data from the World Factbook on the top two primary languages for each country, and the percentage of the population with this as their primary language, was used.</p> <p>The sum of the percentage of the top two languages were scaled to sum to 100, and then each percentage was applied to the count of datasets that include that language. These were then summed to get the total datasets available for each country.</p>

Accessibility of high-value datasets (HVD)	The OECD Open, Useful and Re-usable data (OURdata) Index benchmarks efforts made by governments to design and implement national open government data policies. With subsequent editions released in 2017 and 2019, the Index has remained a valuable resource for policymakers and serves as a key public governance indicator, assessing the progress governments have made in ensuring open data to support policy reform.	Percentage of available high-value datasets that are provided in easily accessible and reusable formats (non-proprietary (open) format, through API, up to date, good metadata quality, on a central portal)
Availability of high-value datasets (HVD)	More details can be found in the OECD reports (Forthcoming ^[27] ; 2026 ^[19]).	Percentage of high value datasets made available by governments (available = open license, machine-readable, free of charge).
Existence of an open data strategy		A score of 1 was assigned to countries saying they have an open data strategy, and 0 otherwise.
Number of supercomputers	The TOP500 supercomputer list is a ranking of the world's most powerful supercomputers, based on their performance in solving linear equations using the High Performance Linpack (HPL) benchmark. The HPL benchmark measures a system's ability to solve dense systems of linear equations, providing a standardised metric for comparison across diverse architectures and configurations. The list serves as a key resource for understanding trends in high-performance computing, highlighting advancements in hardware, software, and computational techniques. The data for the TOP500 are gathered through submissions from manufacturers, users, and other stakeholders, and are validated to ensure accuracy and consistency.	By filtering for geographical location, this indicator sums of all supercomputer's Rmax and Rpeak (separately) from the TOP500 list were calculated for each territory, and then each country was indexed between 0 and 1. The latest data (i.e., from November) were selected for each year. <u>Disclaimer:</u> traditional HPCs are primarily used for scientific computation/modeling and simulation tasks, while AI oriented GPU clouds are used for training large-scale AI models. Data for the latter were however not available, hence the reliance on HPCs.
Strategic approach to cloud infrastructure	The OECD Digital Government Index assesses countries' digital government by looking at the comprehensiveness of strategies and initiatives in place to be able to leverage data and technology to deliver a whole-of-government and human-centric digital transformation of the public sector. The assessment is based on the six dimensions of the OECD Digital Government Policy Framework: 1) digital by design, 2) data-driven public sector, 3) government as a platform, 4) open by default, 5) user-driven, and 6) proactiveness. The DGI is a composite index that takes values from 0 to 1, where 1 indicates the highest digital government maturity and 0 indicates low and/or fragmented progress across organisations. More details can be found in the OECD reports (Forthcoming ^[27] ; 2026 ^[19]).	Countries replied to the question "How is the strategic approach to cloud infrastructure development managed by your central/federal government?" and had to select one option between: A. There is a dedicated strategy/policy on cloud infrastructure for the central/federal government; B. Cloud infrastructure is included as part of the national digital government strategy/policy (NDGS); C. Other; D. None of the above. They were assigned a score of 1 if they replied A, 0.5 points for answer B and 0 points for answers C or D.
Use of AI in the central/federal government		Countries replied to the question "Has the central/federal government used AI to improve the following? Select all that apply among the three, and provide the requested information at least for one initiative per answer option". They were assigned 1/3 of a point per each improved area, for a maximum score of 1 point.
Amount of venture capital investments in AI	Preqin is a private company which collects data on private equity transactions, funds and fund managers. Preqin covers various asset classes including VC investments. Information provided includes the number of venture capital deals, their count and value are included as well as information on the industry relevant to the deal.	Start-ups are identified as AI or data start-ups based on Preqin's cross-industry and vertical categorisation. The sums of the value of all deals within each territory (determined by the company receiving the funding) were calculated. .
Existence of regulatory sandboxes on AI	A regulatory sandbox is a limited form of regulatory waiver or flexibility for firms, enabling them to test new business models with reduced requirements. Such sandboxes are useful for assessing the potential adjustments needed in the legal frameworks to better accommodate the evolving capabilities of AI, thus providing a more informed basis for future legislation and policy decisions.	A score of 1 was assigned to countries with planned strategies regarding regulatory sandboxes in AI, 2 points if these strategies are effectively implemented, and 0 points otherwise. This approach aims to differentiate between countries at different stages of regulatory sandbox implementation to reflect varying degrees of commitment and progress in fostering AI innovation and regulation.

	To compile a directory of AI sandbox initiatives for the indicator under review, the OECD Secretariat relied on the DataSphere Initiative's report " <i>Sandboxes in AI</i> " which serves as the primary source for our dataset.	
Membership to the International Network of AI Safety Institutes	An AI Safety Institute (AISI) is a government-backed institution dedicated to fostering the safe and secure development and deployment of AI. The International Network of AI Safety Institutes was established in November 2024, on behalf of the AI safety institutes and government mandated offices that facilitate AI safety and evaluation from Australia, Canada, the European Commission, France, Japan, Kenya, Korea, Singapore, the United Kingdom, and the United States. The mission statement of the Network is as follows: "Our institutes and offices are technical organisations that aim to advance AI safety, help governments and society understand the risks posed by advanced AI systems, and suggest solutions to address those risks in order to minimize harm. Beyond mitigating risks, these institutes and offices play a crucial role in guiding the responsible development and deployment of AI systems."	In 2023, before the establishment of the Network, manual research was conducted to provide a score of 1 for countries with an established AI Safety Institute and 0 to those without. In 2024 and in future iterations, a score of 1 is given to members of the Network, and 0 for non-members. Common instruments (e.g. EU-level initiatives) are not taken into account. The rationale behind this choice is that including supranational efforts would not accurately reflect which individual countries are taking national leadership and investing directly in the institutional infrastructure needed to ensure AI safety.
Existence of national strategy on AI	The existence of a national AI strategy indicates a government's overarching policy commitment and strategic direction for AI development and adoption. National AI strategies were identified using the OECD.AI Policy Navigator and desk research by OECD analysts.	A value of 1 is given to those countries with a national AI strategy. Future updates to this indicator may take into account the age and last update of the strategy.
Net AI talent migration i.e. net flow migration of AI talent per 10000 people, derived from the self-identified locations of LinkedIn member profiles	LinkedIn is an online platform designed for professional networking and career development. It serves as a virtual hub where individuals can connect with peers, explore job opportunities, and share industry insights. LinkedIn's online platform encompasses millions of user profiles, which include detailed information on professional backgrounds, skills, and endorsements. The dataset covers annual information on skills and jobs of LinkedIn members from 2016 to 2023 across industry and country for data on AI talent, and for 2019 to 2023 for skills migration. Countries are included in the sample if the LinkedIn posts represent 40% of the real work force and they display at least 10 AI hires in any given month on LinkedIn for this country.	The indicator describes the net flow migration of AI talent per 10000 people, for each country per year. A negative (positive) value indicates that a country is experiencing a loss (gain) in its AI talent pool relative to the size of its professional community. Net AI talent migration between country A and country B – for country A – is calculated as follows: $\text{Net AI Talent Migration}_{((a,b,t))} = \frac{\text{Net AI Talent flows}_{((a,b,t))}}{\text{Member count}_{((a,t))}}$ The numerator represents the net flow of AI talent between countries A and B at time t. It is calculated as the difference between the number of AI talents who moved from country B to country A and those who moved from country A to country B during the period from t-1 to t. The denominator represents the total number of LinkedIn members in country A at time t, regardless of their AI skills or jobs. Therefore, the Net AI Talent Migration measure is normalised by the total LinkedIn member count in country A at time t, not specifically by the AI talent pool in country A.
Density of LinkedIn members having explicitly added AI skills to their profile and/or occupied in an AI job		A job profile on LinkedIn is considered an AI talent if they have explicitly added AI skills to their profile and/or they are occupied in an AI job. AI talent concentration at the country level is calculated using the counts of AI talent divided by the counts of total LinkedIn members in that country.
Number of high and very high impact software projects	Following the methodology developed by Gonzalez et al. (2020 ^[28]), OECD.AI identifies the publicly available GitHub projects containing AI code. The methodology uses the GitHub API to curate a list of relevant repository topic	AI Projects are classified based on a set of criteria. The project must have a size greater than zero files and must be accessible via the GitHub API and GHTorrent. It is important that the project is a software project and not a tutorial, homework assignment, coding

	<p>labels related to “artificial intelligence”, “deep learning” and “machine learning” and then searches for projects with these specific labels. Including the three search terms, the result led to 439 topic labels that can be used to identify a GitHub repository as artificial intelligence-related if they matched with at least one of the labels. Overall, 53,427 public repositories were identified based on this methodology, and further classified based on their quality to assess the type of contribution to a given public AI project.</p>	<p>challenge, ‘resource’ storage or collection of model files/code samples. The level of contribution to a given public AI project is measured by the number of “commits” made to it. Further, the last commit must have been made within the year 2019.</p> <p>Two quality measures are used to classify public AI projects:</p> <ol style="list-style-type: none"> 1. Project impact: the impact of an AI projects is given by the number of copies (i.e. “forks”) made of that project. Low impact (0 forks), Medium impact (1-5 forks), High impact (6-100 forks), Very high impact (>100 forks). 2. Project popularity: the popularity of an AI projects is measured by the number of followers (i.e. “stars”) received by that project. Low popularity (0 stars), Medium popularity (1-5 stars), High popularity (6-100 stars), Very high popularity (>100 stars) <p>The indicators are calculated as fractional counts by country. The fractional count represents a country’s share of contributions to a specific AI project. To calculate the fractional count, an AI project is divided equally by the total number of contributions made to it. This indicator uses the sum of high and very high impact software projects.</p>
Average importance of search terms related to artificial intelligence average value (score index)	<p>Google Trends offers access to a largely unfiltered sample of search queries made to Google, which are anonymised and categorised by the topic of the search query. This data is instrumental in measuring public interest in various topics and tracking the evolution of that interest over time, across different languages, and geographic regions.</p> <p>Google Trends provides Search Volume Indices, which measure search intensity by calculating the number of searches for a given keyword relative to the total number of searches in a specific location and period. Users can conduct queries based on keywords, categories of keywords, or topics. Queries based on keywords are language-specific and can be subject to ambiguity. In contrast, categories and topics are harmonised across languages, allowing for comparability across countries. This harmonisation enables the representation of search interest in broader concepts rather than specific terms.</p>	<p>Data cleaning involved removing repetitive or irrelevant related topics. This included topics such as ‘Artificial’, ‘Intelligence’, ‘Machine’, and ‘Learning’, as well as topics spuriously related to acronyms (e.g., ‘Adobe Illustrator’ for ‘AI’ and ‘Neuro-Linguistic Programming’ for ‘NLP’). Semantic similarity was applied to filter out topics that were least related to the original key search topic. The filter, based on cosine similarity, used a threshold of 0.15 to systematically exclude the least related topics to the original search topic.</p>
Number of international research collaborations on AI topics	<p>Scopus is Elsevier’s expertly curated abstract and citation database, with over 75 million indexed records. The data used to construct OECD.AI visualisations includes scholarly articles, conference proceedings, reviews, book chapters and books from a subset of AI-related resources belonging to Elsevier.</p>	<p>OECD.AI calculates research collaborations between different entities, either institutions or countries (Country names and codes in OECD.AI abide by the “<i>OECD Guidelines regarding the use of the list of names of countries and territories</i>”). This is done by assigning each paper to the relevant institutions and countries on the basis of the authors’ institutional affiliations.</p> <p>To avoid double counting, collaborations are considered to be binary: either an entity collaborates on a paper (value=1) or it does not (value=0). The shared paper counts as one toward the number of collaborations between two entities.</p>
Legally binding instruments Voluntary multilateral initiatives	<p>The Policy Navigator is a live database updated regularly by official contact points from countries and international organisations and OECD.AI experts.</p>	<p>A score of 1 was given to countries that signed the declaration/initiative and 0 to other countries that did not sign it.</p>

<p>Non-binding declarations and political commitments (from Policy Navigator)</p>	<p>Users can see who submitted or updated each entry and when. A filter for "International and multi-stakeholder co-operation on AI" was applied to the database.</p> <p>The results were further refined to retain policies that fulfilled two criteria: being AI-specific; and applicable across jurisdictional borders for at least one OECD country. Events and conferences were also excluded from the selection.</p> <p>This resulted in the inclusion of 9 declarations and initiatives on AI. These were further classified into three categories:</p> <p>Legally binding instruments (Committee on Artificial Intelligence), voluntary multilateral initiatives (G7 Hiroshima Process on Generative Artificial Intelligence, and the FAIR Forward initiative), and non-binding declarations and political commitments (the Bletchley declaration, the Seoul declaration, the Montevideo declaration, the Santiago declaration, the declaration on AI in the Nordic-Baltic Region, and the Cartagena de Indias Declaration for Governance).</p>	<p>For voluntary multilateral initiatives and non-binding declarations and political commitments, the commitment of the European Union does not necessarily imply obligatory compliance by its member states. Therefore, individual EU countries that did not explicitly sign the declaration received a score of 0.</p> <p>For legally binding instruments, namely the Committee on Artificial Intelligence, the EU signed on behalf of its member states, and therefore individual EU countries received a score of 1.</p> <p>Each signing was weighted equally, for each declaration or initiative.</p>
<p>National participation in AI standardisation committees</p>	<p>ISO/IEC JTC 1/SC 42 is a committee of the Joint Technical Committee (JTC 1), a collaboration between the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). SC 42 specifically focuses on AI. It is responsible for developing international standards that address the entire AI ecosystem, including, for example, safety considerations or specific AI applications. SC 42 aims to create frameworks and guidelines that ensure interoperability, safety, and ethical development and deployment of AI technologies, fostering global collaboration and consistency in AI development and implementation across industries.</p>	<p>A score of 1 was assigned to countries with Participating (P) member status and 0.5 points to those with Observer (O) member status in ISO/IEC JTC 1/SC 42. This scoring decision reflects the different levels of engagement and influence countries have within the committee. Participating members play a more active role in shaping AI standards through voting rights, regular attendance at meetings, and contributions to working groups. Observer members, while having access to documents and discussions, have a more limited impact on the standardisation process.</p>