Industry Activities

Production and Trade of Knowledge- and Technology-Intensive Industries

Technical Appendix

NSB-2022-6

April 19, 2022

Table of Contents

List of Tables

Technical Appendix

Industry Data, Methodology, and Terminology

This thematic report uses a variety of data sources, including U.S. data on industry value added from the U.S. Bureau of Economic Analysis (BEA), U.S. employment and occupation data from the U.S. Census Bureau's American Community Survey (ACS), international industry production data from IHS Markit, trade data from the Organisation for Economic Co-operation and Development (OECD), and various data sources on artificial intelligence (AI) and biotechnology. This appendix provides information on the classification of industries and the main data sources utilized in this report. Internationally comparable data are generally compiled from multiple national sources and are prone to varying issues of quality and reliability that need caution when making international comparisons. In addition, most data used in this report are periodically revised to reflect new information and methodology improvements.

Classification of Industries Based on Research and Development Intensity

This report defines KTI industries using an OECD taxonomy of economic activities based on research and development (R&D) intensity developed by Galindo-Rueda and Verger (2016). The OECD taxonomy clusters industries into five R&D intensity groups—high, medium-high, medium, medium-low, and low—based on a measure of R&D performance intensity computed as the ratio of each industry's business R&D expenditures to the industry's value-added output.[1] KTI industries include industries in the high and medium-high R&D intensity groups. The full classification of industries is presented in **Table SAKTI-1**.

Table SAKTI-1

OECD classification of industries, by R&D intensity

(List of industries and percent)

R&D intensity category	Manufacturing			Nonmanufacturing		
	ISIC, Rev.4, industry code	Name	R&D intensity (%)	ISIC, Rev.4, industry code	Name	R&D intensity (%)
High R&D intensity	303	Air and spacecraft and related machinery	31.69	72	Scientific research and development	30.39
	21	Pharmaceuticals	27.98	582	Software publishing	28.94
	26	Computer, electronic, and optical products	24.05			
Medium-high R&D intensity	252	Weapons and ammunition	18.87	62 - 63	IT and other information services	5.92
	29	Motor vehicles, trailers, and semi-trailers	15.36			
	325	Medical and dental instruments	9.29			
	28	Machinery and equipment nec	7.89			
	20	Chemicals and chemical products	6.52			
	27	Electrical equipment	6.22			
	30X	Railroad, military vehicles, and transport nec (ISIC 302, 304, and 309)	5.72			
Medium R&D intensity	22	Rubber and plastic products	3.58			
	301	Building of ships and boats	2.99			
	32X	Other manufacturing except medical and dental instruments	2.85			
	23	Other non-metallic mineral products	2.24			
	24	Basic metals	2.07			

Table SAKTI-1

OECD classification of industries, by R&D intensity

(List of industries and percent)

R&D intensity category	Manufacturing			Nonmanufacturing		
	ISIC, Rev.4, industry code	Name	R&D intensity (%)	ISIC, Rev.4, industry code	Name	R&D intensity (%)
	33	Repair and installation of machinery and equipment	1.93			
Medium-low R&D intensity	13	Textiles	1.73	69 - 75X	Professional, scientific and technical activities except scientific R&D services (ISIC 69 - 75 less 72)	1.76
	15	Leather and related products	1.65	61	Telecommunications	1.45
	17	Paper and paper products	1.58	05 - 09	Mining and quarrying	0.8
	10 - 12	Food products, beverages and tobacco	1.44	581	Publishing of books and periodicals	0.57
	14	Wearing apparel	1.40			
	25X	Fabricated metal products except weapons and ammunition (ISIC 25 less 252)	1.19			
	19	Coke and refined petroleum products	1.17			
	31	Furniture	1.17			
	16	Wood and products of wood and cork	0.70			
	18	Printing and reproductions of recorded media	0.67			
Low R&D intensity				64 - 66	Financial and insurance activities	0.38
				35 - 39	Electricity, gas, and water supply; waste management; and remediation	0.35
				59 - 60	Audiovisual and broadcasting activities	0.32
				45 - 47	Wholesale and retail trade	0.28
				01 - 03	Agriculture, forestry, and fishing	0.27
				41 - 43	Construction	0.21
				77 - 82	Administrative and support services	0.18
				90 - 99	Arts, entertainment, repair of household goods, and other services	0.11
				49 - 53	Transportation and storage	0.08
				55 - 56	Accommodation and food service activities	0.02
				68	Real estate activities	0.01

ISIC, Rev.4 = International Standard Industrial Classification, Revision 4; IT = information technology; nec = not elsewhere classified; OECD = Organisation for Economic Co-operation and Development.

Note(s):
R&D intensity is measured as the ratio of global R&D expenditures to global value added output of industry. The global R&D and value added excludes several economies, including China and India, due to incomplete or missing industry value added and R&D data. Industries are classified according to ISIC, Rev.4.

Source(s):
Galindo-Rueda F, Verger F, *OECD Taxonomy of Economic Activities Based on R&D Intensity*, OECD Science, Technology and Industry Working Papers, 2016/04, OECD Publishing, Paris (2016).

Science and Engineering Indicators

Industries are classified according to the United Nations' International Standard Industrial Classification of All Economic Activities, Revision 4 (ISIC, Rev.4). This classification delineates the economic activities of industries based on similarities in inputs and factors of production, the process and technology of production, and characteristics and use of outputs. For economic units that engage in several types of independent activities, considerable proportions of the activities of the unit may be included in more than one class of ISIC. If the unit cannot be split into separate statistical units based on these activities, it is classified based on the principal activity, the activity that contributes most to the value added of the unit.[2]

The R&D-intensity measure captures the R&D directly performed by industries but does not capture any R&D embedded in an industry's purchases of intermediate inputs and capital goods. For any given industry, the R&D intensity is computed as a weighted average of corresponding industry R&D intensities for a core sample of countries using value added in purchasing power parities as weights.[3] Because this taxonomy is based on average R&D intensities across countries, the R&D intensities for industries in individual countries are likely to vary from the average.

The OECD taxonomy is sensitive to the choice of the group of countries. The OECD analysis underlying this classification does not fully capture global production and R&D because of the exclusion of several economies that have incomplete data, including Brazil, India, and China. Because these economies may have large global production shares and R&D intensities that may be substantially different from the average of the core sample of economies included in the OECD analysis, their exclusion may result in differences in classification of R&D intensity.[4]

The R&D intensities are also sensitive to the national accounting conventions used to compile and report the output data. The output data are compiled and reported under the 1993 System of National Accounts (SNA) prior to the capitalization of R&D expenditures with the 2008 SNA. The treatment of R&D expenditures as an investment directly increases the measure of value added and decreases the R&D intensities; the more intensive the industries, the higher the downward impact in the R&D intensities. In a series of robustness checks, the authors estimate that with the 2008 SNA standard, the R&D intensities drop by 20% in the highest group, 6% in the medium-high group, 3% in the medium and medium-low groups, and 1% in the low group. The classification of industries in these groups, however, remains stable.[5]

U.S. Value Added, by Industry and State

The U.S. industry production data come from the Industry and Regional Economic Accounts of BEA, which publishes annual value added by industry statistics for over 130 industries and annual gross output by industry statistics for over 400 industries. These industries are based on the North American Industry Classification System (NAICS)—specifically, the 2012 revision of NAICS. The value-added data discussed in this report are presented in nominal values; the value of production is measured at current market prices, and it is not adjusted for inflation.

For KTI industries, the U.S. data need to be converted from the 2012 NAICS basis to the ISIC Rev.4 basis. The concordance used is presented in **Table SAKTI-2**. For most industries, this adjustment of value added is straightforward. In some cases, however, the value added of a detailed industry, not available in the published data, needs to be estimated. In this case, value added for detailed industries was estimated from corresponding gross output ratios as follows:

$$VA_{detailed\ industry} = \left(\frac{GO_{detailed\ industry}}{GO_{aggregate\ industry}} \right) * VA_{aggregate\ industry}$$

This approach assumes that the ratio of intermediate consumption relative to industry output for the detailed industry is the same as the ratio of the intermediate consumption to the industry output for the aggregate industry.

Table SAKTI-2

2012 NAICS to ISIC, Rev.4, concordance for U.S. value added, by industry data

(List of industries)

R&D intensity category	Value added, by industry		Related 2012 NAICS		ISIC, Rev.4	
	Industry title	Industry code	Industry title	Industry code	Industry title
High R&D intensity	Pharmaceutical and medicine manufacturing	3254	Pharmaceutical and medicine manufacturing	21	Manufacture of pharmaceuticals, medicinal chemical, and botanical products
	Computer and electronic products	334	Computer and electronic product manufacturing	26	Manufacture of computer, electronic, and optical products
	Optical instrument and lens manufacturing[a]	333314	Optical instrument and lens manufacturing		
	Aerospace product and parts manufacturing	3364	Aerospace product and parts manufacturing	303	Manufacture of air and spacecraft and related machinery
	Software publishers	5112	Software publishers	582	Software publishing
	Scientific research and development services	5417	Scientific research and development services	72	Scientific research and development services
Medium-high R&D intensity	Basic chemical manufacturing	3251	Basic chemical manufacturing	20	Manufacture of chemicals and chemical products
	Resin, rubber, and artificial fibers manufacturing	3252	Resin, synthetic rubber, and artificial synthetic fibers and filaments manufacturing		
	Other chemical manufacturing	3253	Pesticide, fertilizer, and other agricultural chemical manufacturing		
		3255	Paint, coating, and adhesive manufacturing		
		3256	Soap, cleaning compound, and toilet preparation manufacturing		
		3259	Other chemical product and preparation manufacturing		
	Electrical equipment, appliances, and components	335	Electrical equipment, appliance, and component manufacturing	27	Manufacture of electrical equipment
	Agricultural implement manufacturing	3331	Agriculture, construction, and mining machinery manufacturing	28	Manufacture of machinery and equipment nec
	Construction machinery manufacturing				
	Mining and oil and gas field machinery manufacturing				
	Other machinery[b]	3332	Industrial machinery manufacturing		
		333316	Photographic and photocopying equipment manufacturing		
		333318	Other commercial and service industry machinery manufacturing		
		3334	Ventilation, heating, air-conditioning, and commercial refrigeration equipment manufacturing		
		333517	Machine tool manufacturing		
		333514	Special tool, die, jig, and fixture manufacturing		
		333515	Cutting tool and machine tool accessory manufacturing		
		333519	Rolling mill and other metalworking machinery manufacturing		
		3336	Engine, turbine, and power transmission equipment manufacturing		

Table SAKTI-2

2012 NAICS to ISIC, Rev.4, concordance for U.S. value added, by industry data

(List of industries)

R&D intensity category	Value added, by industry		Related 2012 NAICS		ISIC, Rev.4	
	Industry title	Industry code	Industry title	Industry code	Industry title
		3339	Other general purpose machinery manufacturing		
	Plumbing fixture fitting and trim manufacturing[a]	332913	Plumbing fixture fitting and trim manufacturing		
	Ball and roller bearing manufacturing[a]	332991	Ball and roller bearing manufacturing		
	Motor vehicle manufacturing	3361	Motor vehicle manufacturing		
	Light truck and utility vehicle manufacturing				
	Heavy duty truck manufacturing				
	Motor vehicle body, trailer, and parts manufacturing[c]	3362	Motor vehicle body and trailer manufacturing	29	Manufacture of motor vehicles, trailers and semi-trailers
		33631	Motor vehicle gasoline engine and engine parts manufacturing		
		33632	Motor vehicle electrical and electronic equipment manufacturing		
		33633	Motor vehicle steering and suspension component (except spring) manufacturing		
		33634	Motor vehicle brake system manufacturing		
		33635	Motor vehicle transmission and power train parts manufacturing		
		33636	Motor vehicle seating and interior trim manufacturing		
		33639	Other motor vehicle parts manufacturing		
	All other transportation equipment manufacturing[d]	3365	Railroad rolling stock manufacturing	30X	Manufacture of railroad, military vehicles and transport nec (ISIC 302, 304, and 309)
		3369	Other transportation equipment manufacturing		
	Medical equipment and supplies manufacturing	3391	Medical equipment and supplies manufacturing	325	Manufacture of medical and dental instruments and supplies
	Computer systems design and related services	5415	Computer systems design and related services	J62 - J63	IT and other information services
	Data processing, internet publishing, and other information services	518	Data processing, hosting, and related services		
		519	Other information services		

ISIC, Rev.4 = International Standard Industrial Classification, Revision 4; IT = information technology; NAICS = North American Industry Classification System; nec = not elsewhere classified.

[a] The value added for these detailed industries has been estimated using gross output ratios. See Technical Appendix for more details.

[b] Other machinery includes Optical instrument and lens manufacturing and Industrial mold manufacturing industries. In the ISIC, Rev.4, classification, Optical lens manufacturing is under the Computer, electronic, and optical products industry, whereas the Industrial mold manufacturing is under Fabricated metal products, except machinery and equipment. Value added for these detailed industries has been estimated using gross output ratios. See Technical Appendix for more details.

[c] Motor vehicle body, trailer, and parts manufacturing incudes Motor vehicle metal stamping. In the ISIC, Rev.4, classification, Motor vehicle metal stamping is under Fabricated metal products, except machinery and equipment. Value added for Motor vehicle metal stamping has been estimated using gross output ratios. See Technical Appendix for more details.

Employment in U.S. KTI Industries

This report uses the 2019 ACS 1-year Public Use Microdata Sample (PUMS) to estimate employment in U.S. KTI industries. The PUMS is a subsample of respondents in the ACS and contains modifications to protect respondent confidentiality in the microdata.[6] These data were pulled from the U.S. Census Bureau's File Transfer Protocol site on 20 December 2020.[7] This report uses information on the respondent's occupation, industry of employment, sex, race or ethnicity, nativity, and citizenship status.

Occupations are grouped into science, technology, engineering, and mathematics (STEM) and non-STEM based on Census occupational codes. (See *Indicators 2022* report "[2022] The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers" for details on how specific Census occupational codes feed into STEM and non-STEM.) STEM workers are further divided into two groups based on educational attainment: STEM workers with a bachelor's degree or above, and STEM workers without a bachelor's degree. This latter group is also known as the skilled technical workforce. The current analysis includes workers in all occupations aged 16−75 years old, except for those in military occupations (Census occupational codes 9800−9830) or currently attending grade school.

Industry employment in the PUMS is reported by a modified version of the NAICS.[8] To identify workers employed in KTI industries in the PUMS, the modified NAICS codes are assigned to ISIC industries. Excluding computer, electronic, and optical products and machinery and equipment, all KTI-related ISIC industries directly align with the modified NAICS in the PUMS. For example, the ISIC industry aircraft aligns with the modified NAICS industries 33641M1 and 33641M2 (**Table SAKTI-3**).

Table SAKTI-3

Concordance for KTI employment data

(List of industries and rate of partial employment)

ISIC, Rev.4	Industry title	NAICS code[a]	Modified NAICS code[b]	Rate of partial employment in NAICS[c]
C303	Aircraft	3364	33641M1, 33641M2	na
C26	Computer, electronic, and optical products	334, 333314	3341, 3345, 334M1, 334M2, pt 3333	na
	Commercial and service industry machinery	333314	pt 3333	0.204821544
C21	Pharmaceuticals	3254	3254	na
M72	Scientific research and development	5417	5417	na
J582	Software	5112	5112	na
C20	Chemicals excluding pharmaceuticals	3251, 3252, 3253, 3255, 3256, 3259	325M, 3252, 3253, 3255, 3256	na
C27	Electrical equipments	335	335M, 3352	na
J62 - J63	IT services	5415, 518, 519	5415, 5182, 51912, 51913, 5191ZM	na

Table SAKTI-3

Concordance for KTI employment data

(List of industries and rate of partial employment)

ISIC, Rev.4	Industry title	NAICS code[a]	Modified NAICS code[b]	Rate of partial employment in NAICS[c]
C28	Machinery and equipment	3331, 3332, 333316, 333318, 3334, 333514, 333515, 333517, 333519, 3336, 3339, 332913, 332991	33311, pt 3333, 333MS, pt 3335, 3336, pt 332MZ	na
	Commercial and service industry machinery	333316, 333318	pt 3333	0.795178456
	Metal working machinery nec	333514, 333515, 333517, 333519	pt 3335	0.760893075
	Miscellaneous fabricated metal products	332913, 332991	pt 332MZ	0.119286701
C29	Motor vehicles	3361, 3362, 33631, 33632, 33633, 33634, 33635, 33636, 33639	336M	na
C30X	Railroad and other transportation	3365, 3369	3365, 3369	na
C325	Medical and dental equipment	3391	3391	na

na = not applicable.

ISIC, Rev.4 = International Standard Industrial Classification, Revision 4; IT = information technology; KTI = knowledge and technology intensive; NAICS = North American Industry Classification System; nec = not elsewhere classified; pt = partial assignment.

[a] The description of NAICS codes in the modified NAICS codes in the Public Use Microdata Sample (PUMS) can be found here: https://www.census.gov/programs-surveys/acs/microdata/documentation.html.

[b] Modified NAICS codes are used in the PUMS. Theses codes are generally the 4-digit NAICS code, but some industries combine multiple 3- or 4-digit NAICS codes, and these modified codes are both alpha and numeric.

[c] The rate of partial employment is calculated from detailed 6-digit NAICS employment estimates from the 2017 Economic Census.

Source(s):
This concordance was developed using the Census Bureau 2012 NAICS to ISIC, Rev.4, concordance available at: https://www.census.gov/naics/. More information on 2012 NAICS classification is available at: https://www.census.gov/naics/. More information on ISIC, Rev.4, is available at: https://unstats.un.org/unsd/classifications/Econ/Structure.

Science and Engineering Indicators

Two ISIC industries align to multiple modified NAICS industries because some of the modified NAICS industries are too coarse to allow for direct matches.

- **Computer, electronic, and optical products, ISIC industry (C26)**. This ISIC industry concords with the modified NAICS industries 3341, 3345, 334M1, and 334M2 and with part of 3333. Part of the modified NAICS industry 3333, commercial and service industry machinery manufacturing, is assigned to ISIC industry C26, and another part of it is assigned to C28. (More details about this ISIC industry appear below.) The NAICS 4-digit industry 3333 includes NAICS 6-digit industries 333314 (optical instrument and lens manufacturing), 333316 (photographic and photocopying equipment manufacturing) and 333318 (other commercial and service industry machinery manufacturing). The NAICS industry 333314 is part of ISIC C26, and the other NAICS industries are part of C28. This partial assignment is denoted with a "pt" designation in the modified NAICS code column and the NAICS column in **Table SAKTI-3**.

- **Machinery and equipment.** This ISIC industry concords with modified NAICS industries 331, 3332, 333MS, and 3336 and with parts of NAICS 3333, 3335, and 332MZ. **Table SAKTI-3** shows the partial assignments of modified NAICS industries 3333, 3335, and 332MZ.

Employment information from the 2017 Economic Census was used to prorate estimated employment based on the PUMS for modified NAICS industries partially assigned to multiple ISIC industries (i.e., those denoted with "pt" in **Table SAKTI-3**). For example, the 2017 Economic Census reported that about 20% of employment in NAICS industry 3333 (commercial and service industry machinery manufacturing) was 6-digit NAICS industry 333314 (optical instrument and lens manufacturing). About 80% of NAICS industry 3333 was the other 6-digit NAICS industries (333316 and 333318). Hence, 20% of estimated employment for modified NAICS industry 3333 based on the PUMS is assigned to ISIC industry C26, and 80% to ISIC industry C28. Similar calculations were applied to estimated employment based on PUMS for modified NAICS industries 3335 and 332MZ.

IHS Markit Comparative Industry Service

The international industry production data are drawn from a proprietary database compiled by IHS Markit. The IHS Comparative Industry Service (CIS) Forecast Database provides consistent coverage for over 70 countries and over 100 industrial sectors for several macroeconomic indicators, including the value-added data presented in this report. The primary data sources on industry output are the National Income Accounts from the countries' national statistical agencies and also the cross-national organizations, including the Industrial Structure Statistics from the OECD Structural Analysis (STAN) database, the International Yearbook of Industrial Statistics from the United Nations Industrial Development Organization (UNIDO), and the National Accounts Statistics from the United Nations System of National Accounts (UNSNA).

IHS Markit compiles the data in the CIS database using a tiered approach, where data from OECD, UNSNA, and UNIDO form the foundation for all sectors and most countries. These are harmonized data that provide consistency and comparability across countries and across time. The OECD STAN database provides data for the OECD member countries. For countries not included in the OECD database, and for industries whose coverage in STAN is not sufficiently detailed, IHS Markit combines the OECD data with the UNSNA and UNIDO databases. The UNSNA contains national accounts data for most countries in the world and facilitates macroeconomic comparisons among national economies at the broad industry level. Data availability, however, varies across countries and fiscal years because not all UN member countries are able to provide a complete set of data. The UNIDO data set provides highly disaggregated data for the manufacturing sector. Both the UN and OECD data are further supplemented by other international sources such as the International Labor Organization and Eurostat and by individual country sources.

For some countries or economies, IHS Markit collects more granular and timely data directly from their national statistical agencies. These countries or economies include the United States (BEA), Brazil (Statistics Brazil), China (China National Bureau of Statistics), and Taiwan (Taiwan Directorate-General of Budget, Accounting and Statistics; National Statistics). Finally, IHS Markit brings the collected data forward in time as needed using data from individual country sources, global trade associations, and other sources to provide timely measures of industry-level business activity.

The value-added data discussed in this report are presented in current dollars. For countries outside the United States, value added is recorded in the local currency and converted at the prevailing nominal market exchange rate. This choice comes with some limitations, particularly when an economy's currency exchange rates are not market determined.

Trade in Value-Added Database

The trade in value-added indicators come from the 2021 OECD Trade in Value Added (TiVA) database. The OECD TiVA statistics are derived from OECD's Inter-Country Input-Output (ICIO) tables, which provide a globally balanced view of inter-country, inter-industry flows of intermediate and final goods and services.[9] The ICIO tables are constructed using national annual Supply and Use Tables (SUTs) or, if necessary, benchmark Input-Output Tables (IOTs), harmonized to a common format and common industry list, linked with balanced bilateral trade in goods and services, and constrained under

countries' latest SNA main aggregates and industry output and value-added time series. The SUTs and IOTs are sourced from national statistical agencies and harmonized by OECD. They show production of output by all sectors and the allocation of domestic output among intermediate and final uses, including exports. The linkage of input-output data with trade data captures the bilateral exchanges of intermediate goods and services.

For countries in the TiVA database, the overall trade balance is consistent with official national accounts figures. However, the bilateral trade estimates may differ from those reported by national statistical agencies. By necessity, asymmetries in reported bilateral trade (exports of a product reported by country A to country B may differ from country B's reported imports from country A) need to be balanced. There are many reasons for asymmetries, including differences across countries in treatment of various aspects of trade such as re-exports and transit trade. The international statistics community continues its work to improve consistency in measuring international trade flows, particularly in services, where there are substantial differences across national statistics.[10]

The most recent update (2021) of the TiVA database covers 66 economies, including the OECD countries, the European Union (EU) and the G20 countries, and several East and Southeast Asian economies and South American countries for the years 1995–2018.[11] The input-output tables from which the TiVA indicators are derived are based on the 2008 SNA. Indicators are available for 45 industries within a hierarchy based on ISIC, Rev.4.

Along with its advantages, trade in value added presents some measurement challenges. The value added of companies with diversified businesses may be assigned to the single industry that accounts for the largest share of the company's business. A company classified as manufacturing may include services (and vice versa). Disentangling the domestic and foreign content in global value chains is further complicated by fragmentation of production within multinational enterprises and trade in inputs that are further traded as more processed inputs. Quality and availability of data (e.g., SUTs, IOTs, and SNA) in many countries, and inconsistencies in national statistics within countries, are major challenges when constructing global IOTs.

PitchBook Venture Capital Data

The venture capital data shown in the report are from PitchBook Data, Inc., a private-sector financial services company that collects financial and business data on the Web and provides subscription-based data (**https://pitchbook.com/**). PitchBook classifies companies by industry and industry vertical: an industry is a "broad group of companies that operate in the same general space," whereas "an industry vertical is more specific and describes a group of companies that focus on a shared niche or specialized market spanning multiple industries" (PitchBook 2021). AI is defined as an industry vertical, and biotechnology is defined as an industry. Companies are classified in one primary industry but can also be classified in secondary industries. Industry verticals are listed in Table SINV-100 in the *Indicators 2022* report "[2022] Invention, Knowledge Transfer, and Innovation"—where the data are used more extensively than in this report.

Within PitchBook's proprietary industry classification system, biotechnology is an industry within the health care sector and the pharmaceuticals and biotechnology industry group. It is the broad area of biology, involving living systems and organisms (e.g., bacteria, mammalian cells, T cells) to develop or make products. Companies in this category are researching and using biological systems to develop new drugs and therapies for medical patients. Pharmaceutical companies, which are within the same industry group, differ from biotechnology companies in that these companies are primarily involved in manufacturing and distributing drugs, generally from chemical and synthetic processes. Drug discovery is another industry within this industry group, and these companies research and develop new drugs. Many of the companies classified in the pharmaceutical and drug discovery primary industries are also classified in the biotechnology secondary industry. Therefore, this analysis includes companies that are listed in the biotechnology primary industry and secondary industry.

AI is defined for PitchBook investment reporting as companies developing technologies that enable computers to autonomously learn, deduce, and act through utilization of large data sets. The technology enables development of systems that collect and store massive amounts of data and analyze that content to make decisions based on probability and statistical analysis. Applications for AI and machine learning include speech recognition, computer vision, robotic control, and accelerating processes in the empirical sciences where large data sets are essential, such as gene sequencing in life sciences (PitchBook/NVCA 2019).

For the biotechnology industry and AI industry vertical, the analysis includes all completed venture capital deals at all stages (i.e., pre/accelerator/incubator, angel, seed, early stage venture capital, later-stage venture capital, other stages), all rounds, and all series (i.e., seed, series A–series D, and later series). All investor types are included—except government and sovereign wealth fund because these are not considered private investors.

AI and Biotechnology Patents

AI and biotechnology patent data are from OECD Science, Technology and Patents (2022), United States Patent and Trademark Office (USPTO) (2022), and *2022 Indicators* report "[2022] Invention, Knowledge Transfer, and Innovation."

OECD collects data on patent families for the five largest patent offices in the world (IP5 patent families)—European Patent Office, Japan Patent Office, Korean Intellectual Property Office, State Intellectual Property Office of the People's Republic of China, and USPTO. IP5 patent families refer to patents that have been filed in at least two IP offices worldwide, one of which is among the IP5. This avoids double counting of patents filed in multiple jurisdictions. (For more details, see *Indicators 2020* report "[2020] Invention, Knowledge Transfer, and Innovation.") This report uses OECD's definitions of AI-related patents (OECD 2017) and biotechnology patents (OECD 2021).

This report also presents data on patents granted by USPTO. Data on AI patents granted by USPTO are reported in Toole et al. (2020). USPTO patents on biotechnology are reported in Table SINV-59 in the *Indicators 2022* report "[2022] Invention, Knowledge Transfer, and Innovation." These data use the World Intellectual Property Organization definition of biotechnology (Schmoch 2008), which differs from the OECD definition.

Emsi Burning Glass Job Postings Data

The proprietary Emsi Burning Glass (Emsi) data set includes over 30,000 skills from millions of online sources, including job posting websites and resumes. Emsi retrieves information from over 100,000 websites, including company career sites, national and local job boards, and job posting aggregators. More information on Emsi's methodology for job postings can be found on Emsi's website (**https://kb.emsidata.com/methodology/job-posting-analytics-documentation/**).

For this analysis, skills related to biotechnology and AI were entered into the Emsi searchable database to extract job postings that featured one of the highlighted skills. The specific skills used for both biotechnology and AI will be highlighted in the methodology below.

It is important to note that the number of job postings do not reflect the actual number of hires. Job postings, regardless of industry, might not yield an actual person being hired. It is also possible that one job posting may be used to fill several vacancies at a company. Further, some websites may post duplicate postings of a job. To reduce duplicate postings for this analysis, Emsi data were filtered to remove postings that originated from staffing companies. This thematic report uses analysis data from October 2016 to September 2021.

Using the Emsi Application Programming Interface, the top 25 skills were extracted for both AI and biotechnology. These skills groupings are based on a review of job postings collected by Emsi, and they only include skills categorized by Emsi as a hard skill. More information on Emsi's skill clustering methodology can be found on Emsi's website (**https://skills.emsidata.com/faqs#how-are-skills-selected**).

The top 25 AI-related skills identified by Emsi are algorithms, Apache Spark, artificial neural networks, big data, blockchain, computer vision, data science, deep learning, distribute computing, Internet of Things, machine learning, machine learning algorithms, mathematical modeling, natural language processing, Pandas (Python package), predictive analytics, predictive modeling, R (programming language), robotic process automation, Scala (programming language), speech recognition, statistical modeling, TensorFlow, time series, and unstructured data.

The top 25 biotechnology-related skills identified by Emsi are biopharmaceuticals, biostatistics, case report forms, clinical pharmacy, clinical research, clinical study design, clinical trial management systems, clinical trials, drug development, drug discovery, electronic data capture, good clinical practices, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use Guidelines, key opinion leader development, life sciences, medical affairs, medical devices, medical guidelines, nondisclosure agreements (intellectual property law), pharmaceuticals, pharmacovigilance, pre-clinical development, regulatory filings, scientific literature, and Title 21 of the Code of Federal Regulations.

References

Galindo-Rueda F, Verger F. 2016. *OECD Taxonomy of Economic Activities Based on R&D Intensity*. OECD Science, Technology and Industry Working Papers No. 2016/04. Paris: OECD Publishing. Available at **https://doi.org/ 10.1787/5jlv73sqqp8r-en**. Accessed 18 April 2019.

Organisation for Economic Co-operation and Development (OECD). 2017. *OECD Science, Technology and Industry Scoreboard 2017: The Digital Transformation.* Paris: OECD Publishing. Available at **https://www.oecd.org/sti/oecd-science-technology-and-industry-scoreboard-20725345.htm**. Accessed 11 November 2022.

Organisation for Economic Co-operation and Development (OECD). 2021. Methodological Work and Publications—Patents in Selected Fields: Biotechnology. Available at **https://www.oecd.org/sti/inno/intellectual-property-statistics-and-analysis.htm#method**. Accessed 18 November 2021.

Organisation for Economic Co-operation and Development, World Trade Organization (OECD/WTO). N.d. *Trade in Value-Added: Concepts, Methodologies, and Challenges.* Joint OECD-WTO Note. Available at **https://www.oecd.org/sti/ind/ 49894138.pdf**. Accessed 27 June 2019.

PitchBook, National Venture Capital Association (PitchBook/NVCA). 2019. *Venture Monitor 4Q 2018*. Seattle, WA: PitchBook Data, Inc. Available at **https://pitchbook.com/news/reports/4q-2018-pitchbook-nvca-venture-monitor**. Accessed 14 March 2019.

PitchBook. 2021. "What Are Industry Verticals?" Available at **https://pitchbook.com/what-are-industry-verticals**. Accessed 29 September 2021.

Schmoch U. 2008. *Concept of a Technology Classification for Country Comparisons: Final Report to the World Intellectual Property Organisation (WIPO)*. Available at **https://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/ wipo_ipc_technology.pdf**. Accessed 18 November 2021.

Toole A, Pairolero N, Giczy A, Forman J, Pulliam C, Such M, Chaki K, Orange D, Homescu A, Frumkin J, Chen Y, Gonzales V, Hannon, C, Melnick S, Nilsson E, Rifkin B. 2020. *Inventing AI: Tracing the Diffusion of Artificial Intelligence with U.S. Patents*. Alexandria, VA: U.S. Patent and Trademark Office. Available at **https://www.uspto.gov/about-us/news-updates/new-benchmark-uspto-study-finds-artificial-intelligence-us-patents-rose-more**. Accessed 15 October 2021.

Notes

1 Value added is a net measure of output; it is the difference between the value of goods and services produced by an industry (gross output) and the total cost of inputs that were used in production, including energy, materials, and services purchased from other businesses. Value added is used in the computation of R&D intensity instead of gross output because it avoids double counting of intermediate production.

2 More information on ISIC, Rev.4, is available at https://unstats.un.org/unsd/classifications/Econ/Structure.

3 The countries are 27 OECD countries and 2 partner countries or economies (Singapore and Taiwan). The OECD countries are Australia, Austria, Belgium, Canada, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, South Korea, Mexico, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, the United Kingdom, and the United States.

4 The study reports that including partial data from China decreases the average R&D intensity in more than two-thirds of the industries. This is consistent with China producing a large share of global output for these industries while investing a below-sample-average share in R&D. This is particularly the case for computer, electronic, and optical products; transport equipment; and chemicals and pharmaceuticals. The inclusion of China would downgrade, for instance, the computer, electronic, and optical products industry from the high R&D intensive group to the medium-high R&D intensive group. For the impact on other industries, see Galindo-Rueda and Verger (2016:25).

5 See Galindo-Rueda and Verger (2016) for a full discussion of robustness checks.

6 These modifications can include, for example, the "top-coding" of continuous variables with outlying values (such as income or transportation time) and the reduction of category availability for variables with hundreds of categories (such as ancestry or birthplace).

7 See https://www2.census.gov/programs-surveys/acs/data/pums/2019/1-Year/.

8 NAICS codes are all numeric. In PUMS, some NAICS codes are combined into modified NAICS codes, which are identified as being both numeric and alpha. (See how PUMS combines NAICS codes in its modified version of NAICS in "2019 ACS 1-year PUMS Code Lists" at https://www.census.gov/programs-surveys/acs/microdata/documentation.html.)

9 The OECD TiVA database is available at https://www.oecd.org/industry/ind/measuring-trade-in-value-added.htm. The ICIO tables are available at https://oe.cd/icio.

10 For a more detailed description of concepts and methodologies related to the TiVA statistics and a discussion of ongoing projects and initiatives, see OECD–World Trade Organization (n.d.).

11 The G20 is a group of 19 individual countries and the EU. The countries are Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, United Kingdom, and United States.