Preliminary work – please do not circulate

WHAT'S AT THE CORE OF AI?

EXPLORING THE CHARACTERISTICS OF AI-RELATED CORE TECHNOLOGIES

Lea Samek
OECD

Based on ongoing work with Flavio Calvino, Chiara Criscuolo and Hélène Dernis

6th Annual IP Data & Research Conference
31st May 2023

BETTER POLICIES FOR BETTER LIVES

Setting the scene

- **Artificial Intelligence** (AI) is often referred to as a **General Purpose Technology** (GPT) that has the potential to **improve economic outcomes** and tackle societal challenges
(Cockburn, Henderson and Stern, 2018; Brynjolfsson, Rock and Syverson, 2021)

- To better understand the impact of AI on the economy and society as a whole, it is decisive to have a **robust definition of what AI is**, taking into account its complex nature.
  - Technological developments in AI become increasingly **embedded in multiple sectors** of the economy, a characteristic common to GPTs.
  - Hence, **distinguishing AI core technologies from related applications** becomes even more **difficult** and operational definition is still missing.

Patents can provide further insights

… by presenting a **quantitative** and forward-looking indicator of technological developments.

- Pre-selection of AI-related **technological classes** (Fujii and Managi, 2018)
  - Help identifying inventions according to the technical content of underlying patents offering insights into the **qualitative** dimension of technological change (Jaffe and de Rassenfosse, 2017; Hall, Jaffe and Trajtenberg, 2001).
  - But existing classification schemes may not suffice if GPTs, and especially AI, are connected with one another in a network of related technologies by underlying principles and mutual dependencies (Hall and Trajtenberg, n.d.; Petralia, 2020).
- **AI related keywords** to identify AI-related inventions (De Prato and Cardona, 2019; European Commission. Joint Research Centre, 2018)
- **Combination** of the two approaches (Cockburn, Henderson and Stern, 2018; WIPO, 2019; Iori, Martinelli and Mina, 2021)

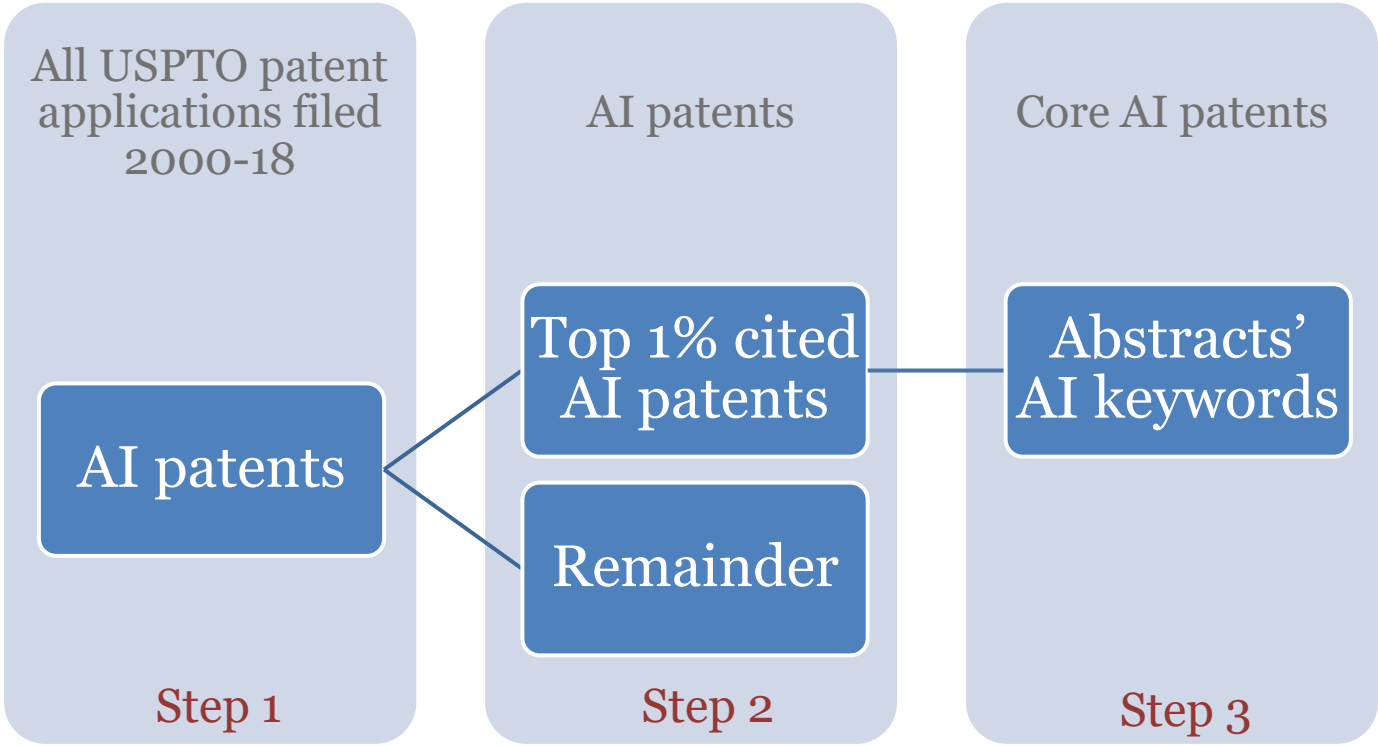
- Exploiting advanced **AI ML techniques** to further identify AI-related innovation by parsing patent text (Giczy, Pairolero and Toole, 2022; Damioli, Van Roy and Vertesy, 2021; Baruffaldi et al., 2020).
  - Baruffaldi et al. (2020) propose an operational definition of AI based on the identification and measurement of **AI-related developments** in **science**, **algorithms** and **technologies**, thereby going beyond the simple assignment of technological fields.

Methodology

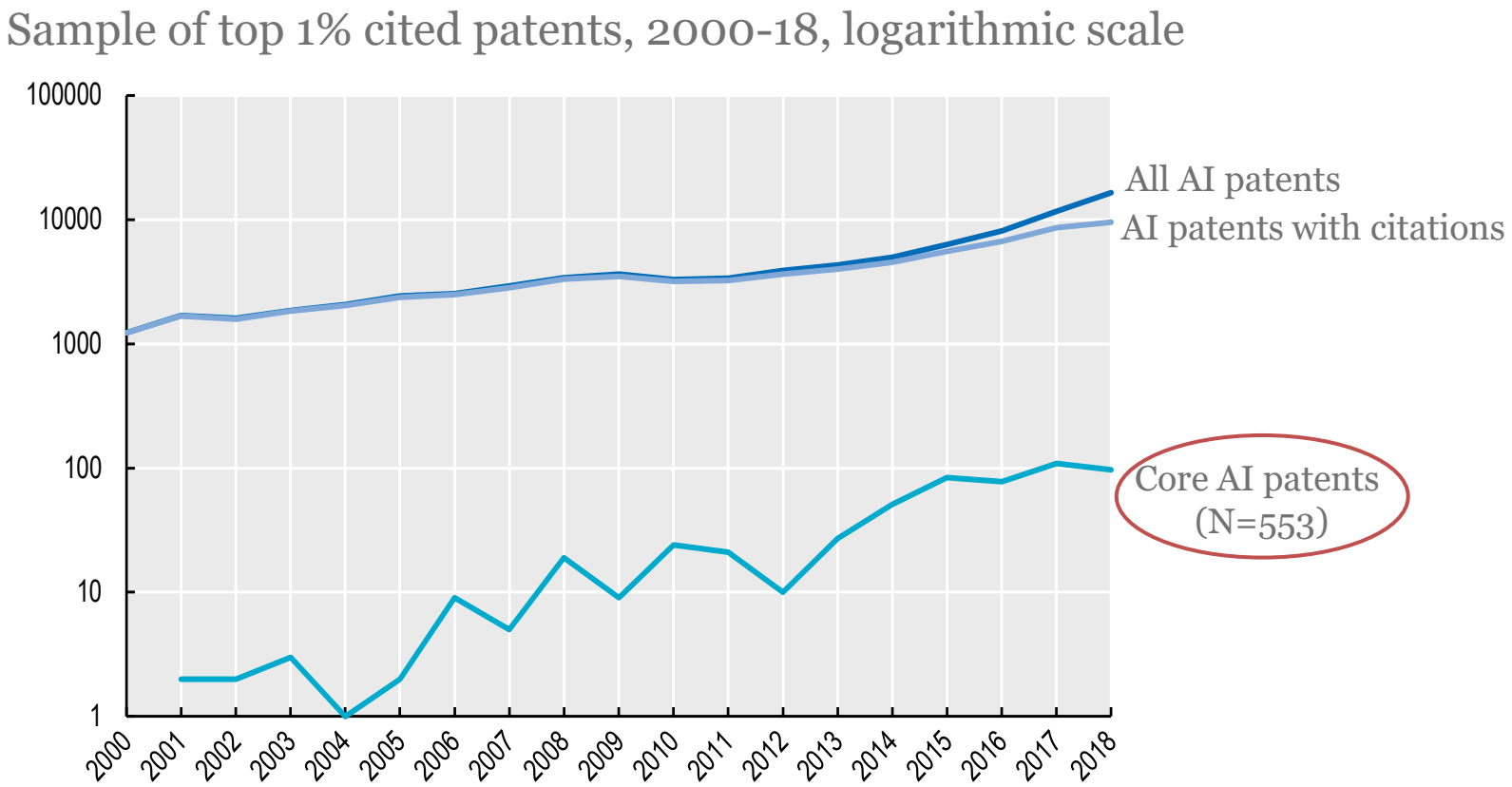
1. Identifying AI-related patents based on IPC/CPC codes closely related to AI and/or AI keyword(s) following Baruffaldi et al. (2020)

2. Using counts of AI-related forward citations (normalised by the average number of forward citations received by patents from a reference cohort) and defining top 1% cited AI patents as “core AI”

3. Exploring AI-related scientific and technological developments in those “core AI” patents based on their abstracts’ AI keywords

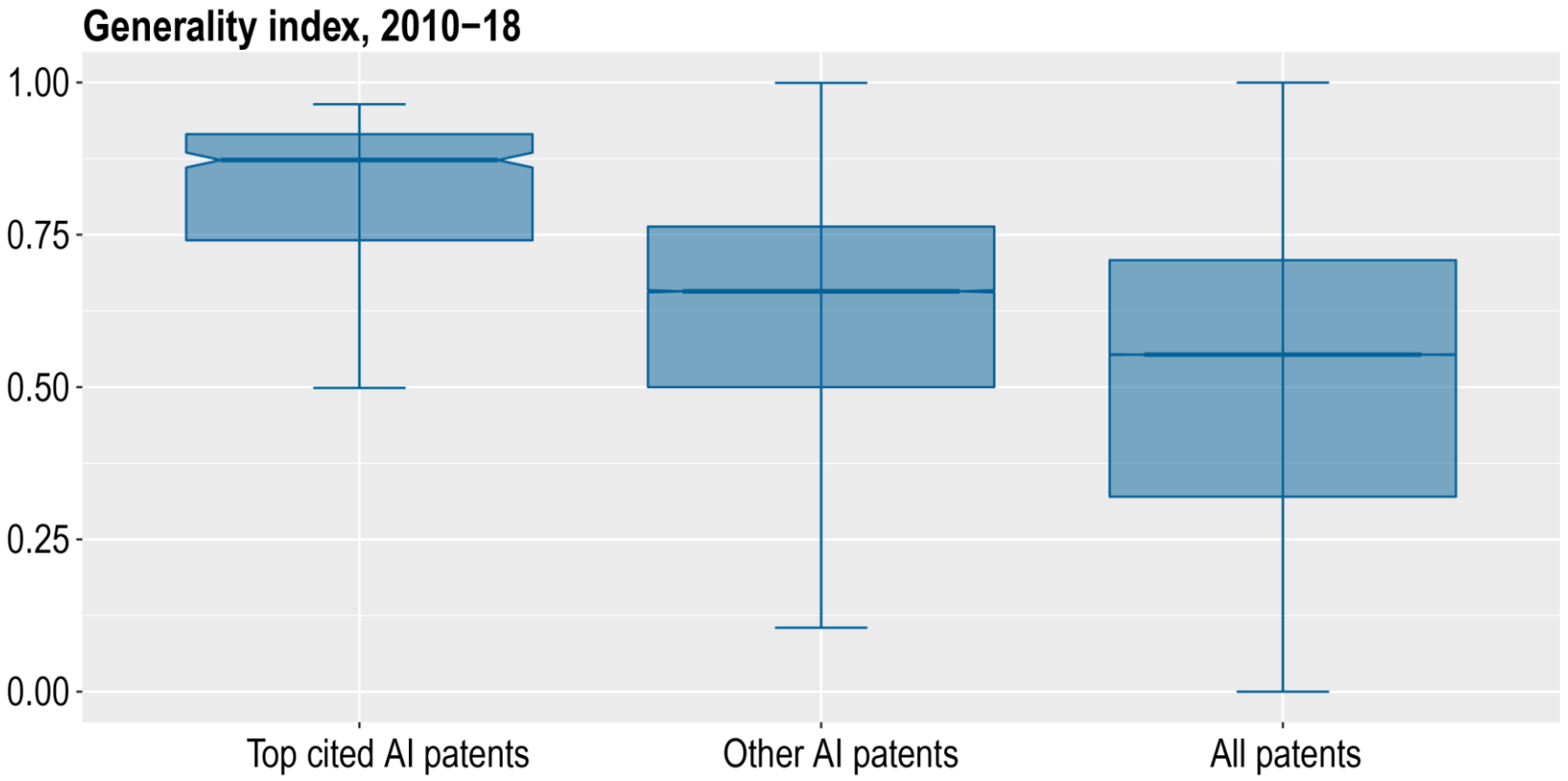


Acceleration of (core) AI patent activity



Note: Data refer to patent applications filed at the USPTO (patents issued and pre-grant publications) that refer to AI technologies. The sample of top 1% cited patents relies on normalised counts of forward citations.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

Patents citing core AI patents belong to a wide(r) range of fields…

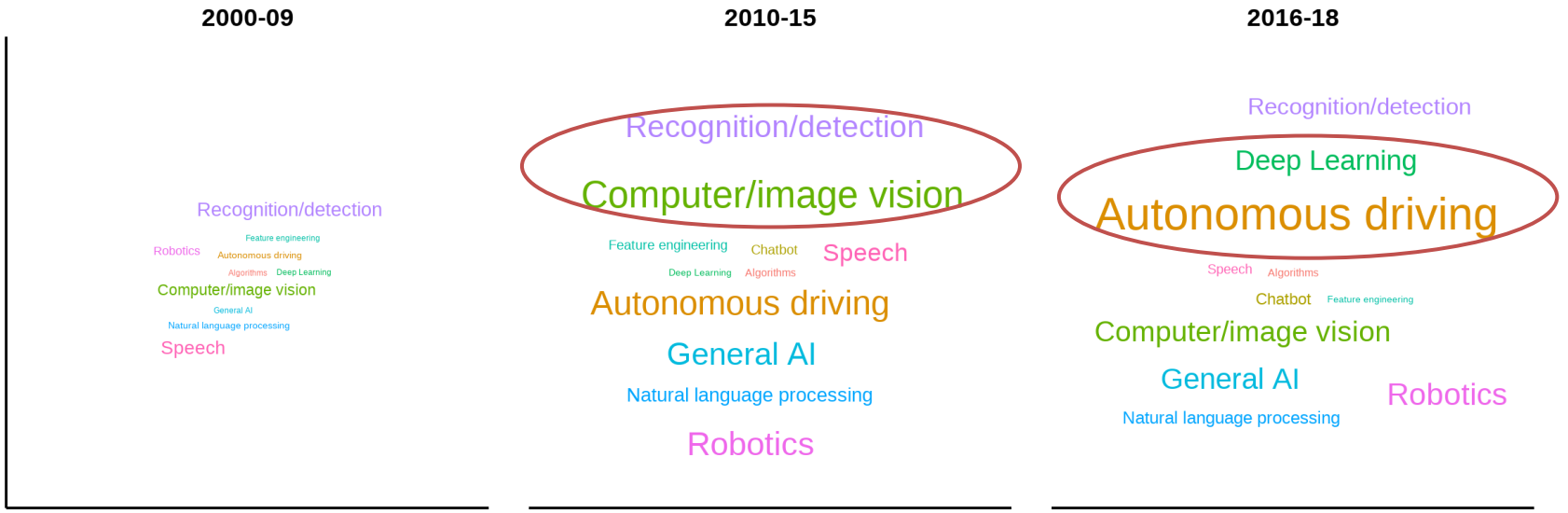


- …Supporting (core) AI's **GPT nature**.
- While overall patents became less general over time, core AI patents **became more general**.
- Core AI patents are also **broader in scope** and rely on a broader **range of technology fields**
  - Although they used to be technologically more radical.

Note: The AI patent sample (up to 553 patents) refers to the top 1% cited patent applications (normalised counts of forward citations) filed at the USPTO (patents issued and pre-grant publications) that refer to AI technologies. The generality index is constructed based on Squicciarini, Dernis and Criscuolo (2013).
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

Exploring topics at the core of AI

AI topics of top 1% cited patents, 2000-18

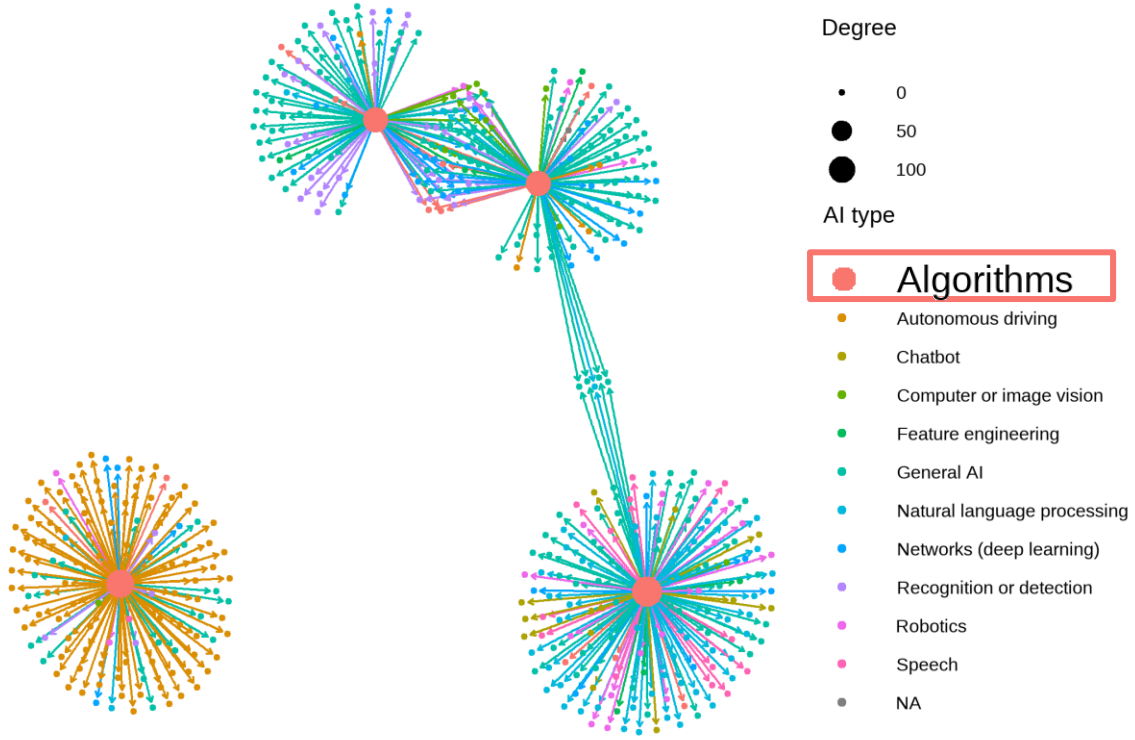


Note: The size of the words reflects the frequency with which each topic appears in the sample, which refers to the top 1% cited patent applications (normalised counts of forward citations) filed at the USPTO (patents issued and pre-grant publications) that refer to AI technologies. An AI core patent can have more multiple AI keywords and hence multiple topics. The analysis uses on a sub-sample (485 patents) of core AI patents with AI topics being imputed based on patents with similar IPC characteristics. Core AI patents without any topics (even after the imputation exercise) are dropped.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

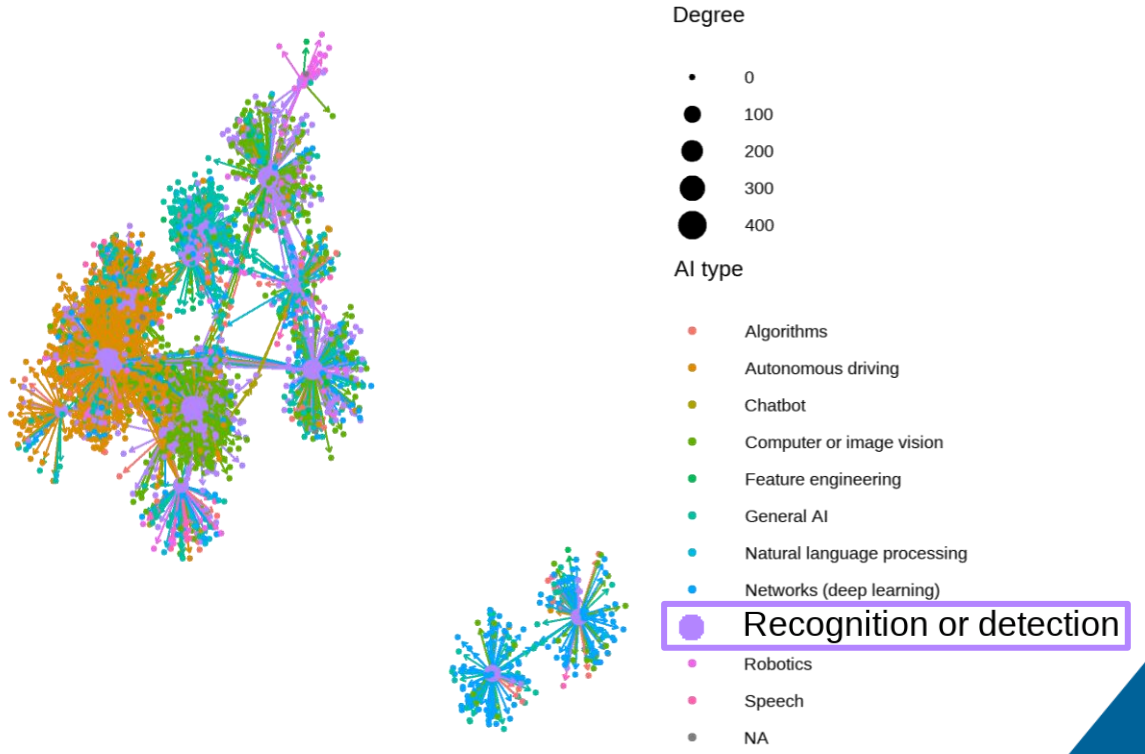
- **General AI, robotics, computer/image vision** and **recognition/detection** related technologies are **consistently** at the core of AI development…
- But more recently **autonomous driving** and **deep learning** gain significant **importance**.
- Technologies related to **feature engineering**, **chatbots**, **algorithms**, **NLP** and **speech** appear **less prevalent**.

While most core AI technologies tend to build on one another…

Algorithms
core AI patents with AI citations, 2016-18



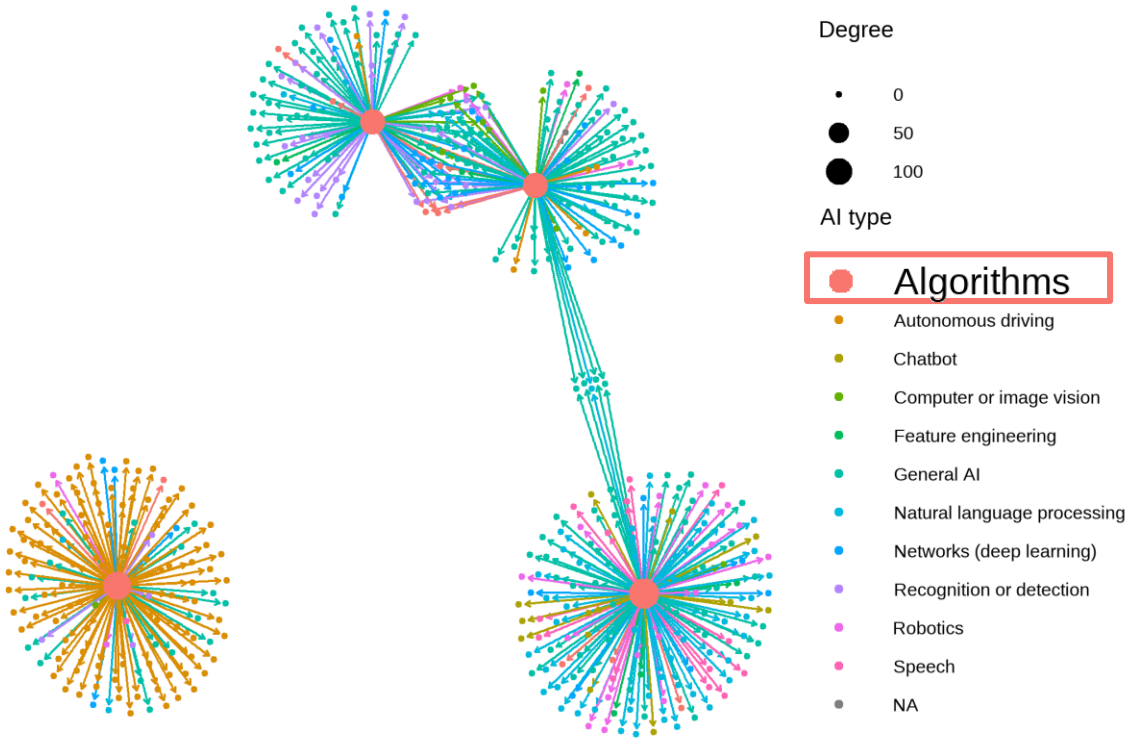
Recognition/detection
core AI patents with AI citations, 2016-18



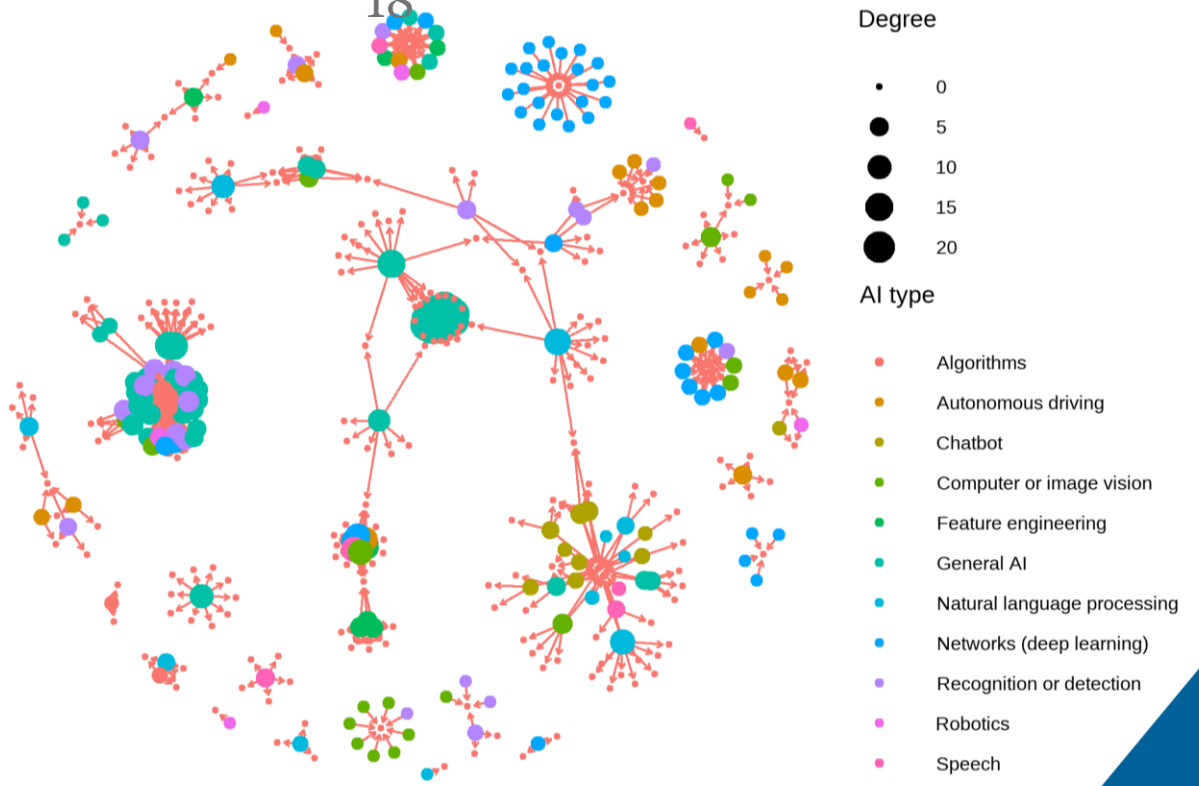
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

While most core AI technologies tend to build on one another…

Algorithms
core AI patents with AI citations, 2016-18



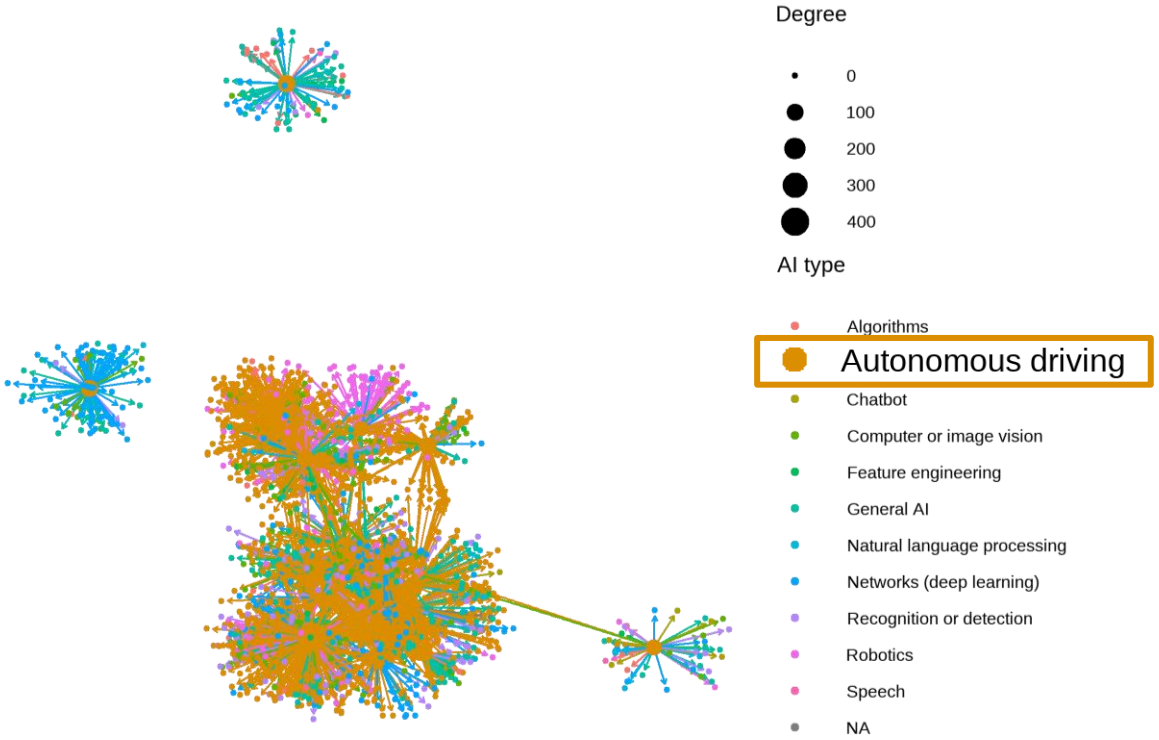
Core AI patents with **algorithm AI citations**, 2016-18



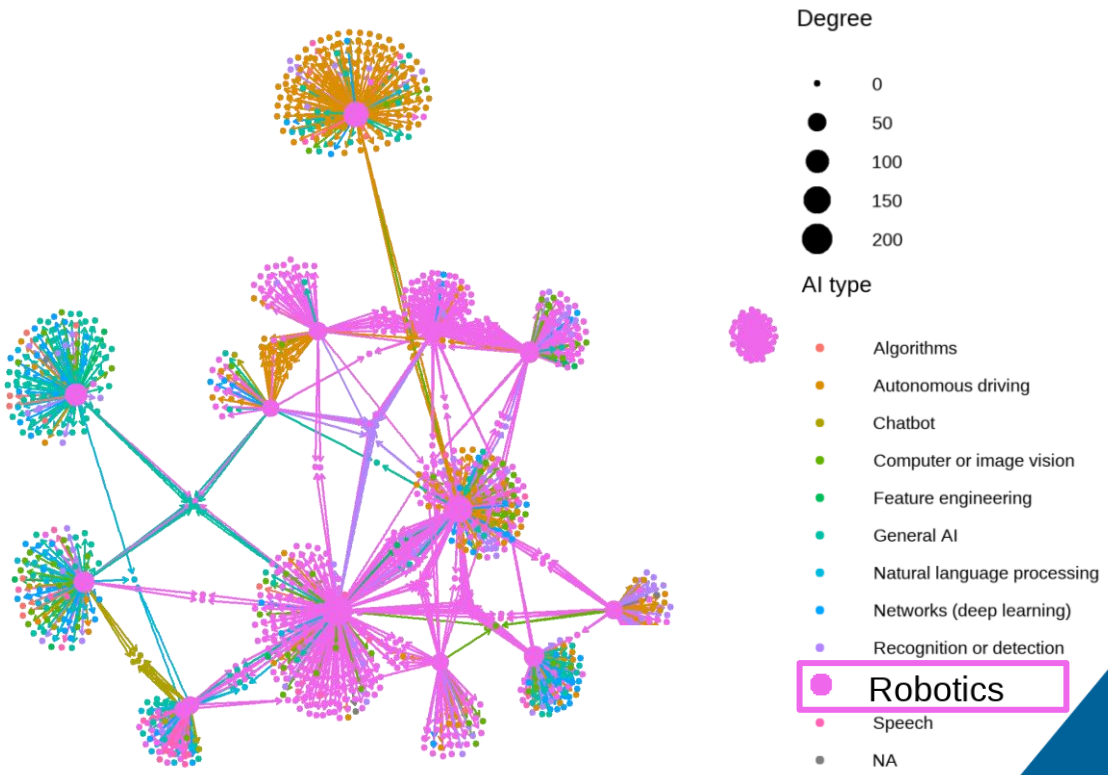
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

… some AI-core applications seem to contribute to their own development

Autonomous Driving
core AI patents with AI citations, 2016-18



Robotics
core AI patents with AI citations, 2016-18



Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

Key takeaways so far

- The initial insights help to better understand the **technological base underlying** the most recent stages of the **digital transformation**.

- Preliminary findings suggest that overall **core AI technologies** appear relevant for a wide number of **later AI innovations** and, at the same time, also seem to rely on **past core AI technologies**.
  - They tend to have a **more general** nature – in terms of e.g., generality and scope – than other technologies, and **used to be more radical** relative to other patents in the earlier stages of development.
  - They increasingly tend to **build on one another**, although some of them – likely applications – appear to increasingly **contribute to their own developments**.

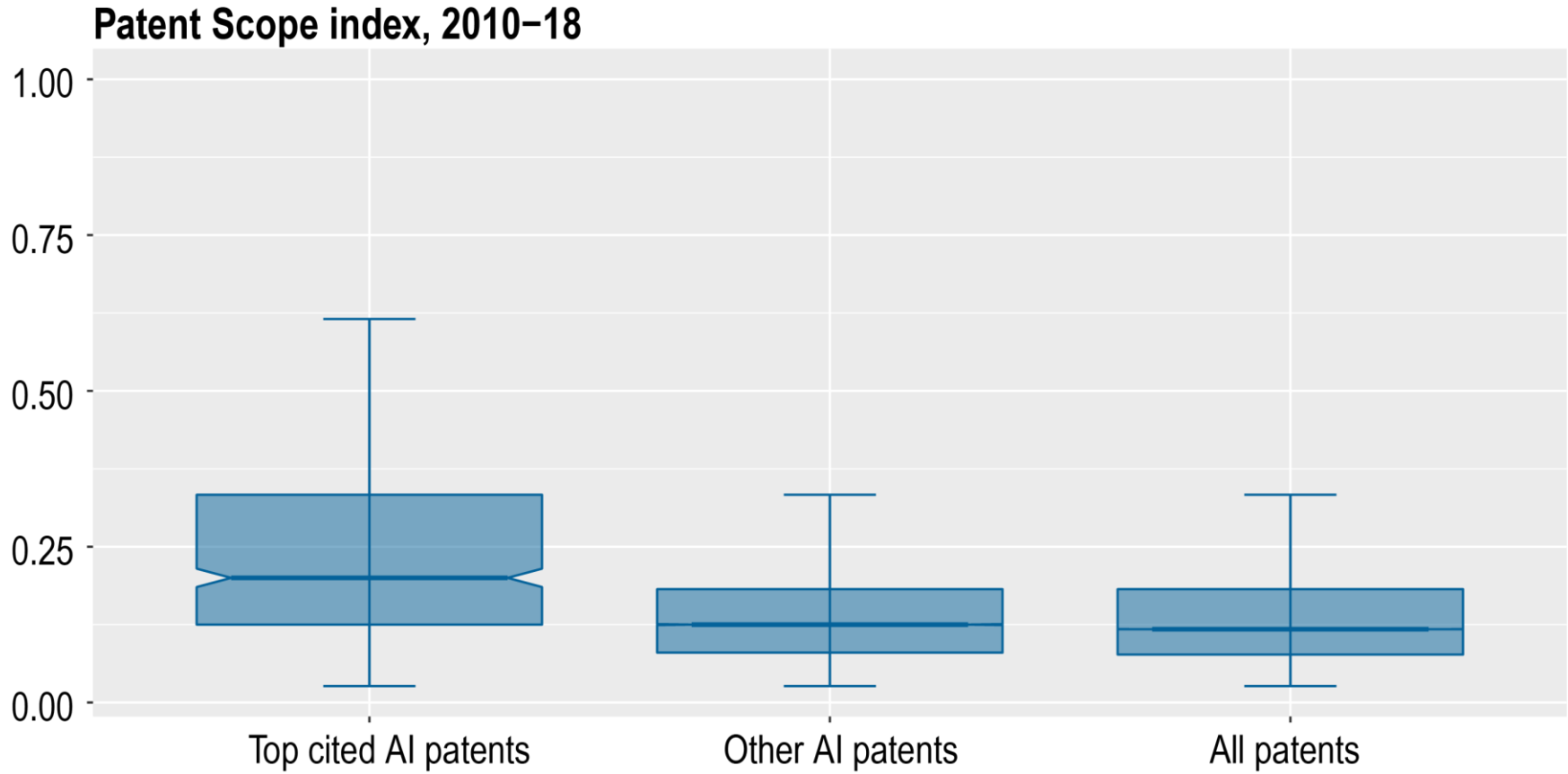
Thank you!

Lea.SAMEK@OECD.org

Additional material

Core AI patents are broader in scope

- Compared to other (AI) patents, core (i.e. top-cited) AI patents are **broader in scope**, especially in recent years, and hence have potentially a **higher technological** and **market value**.

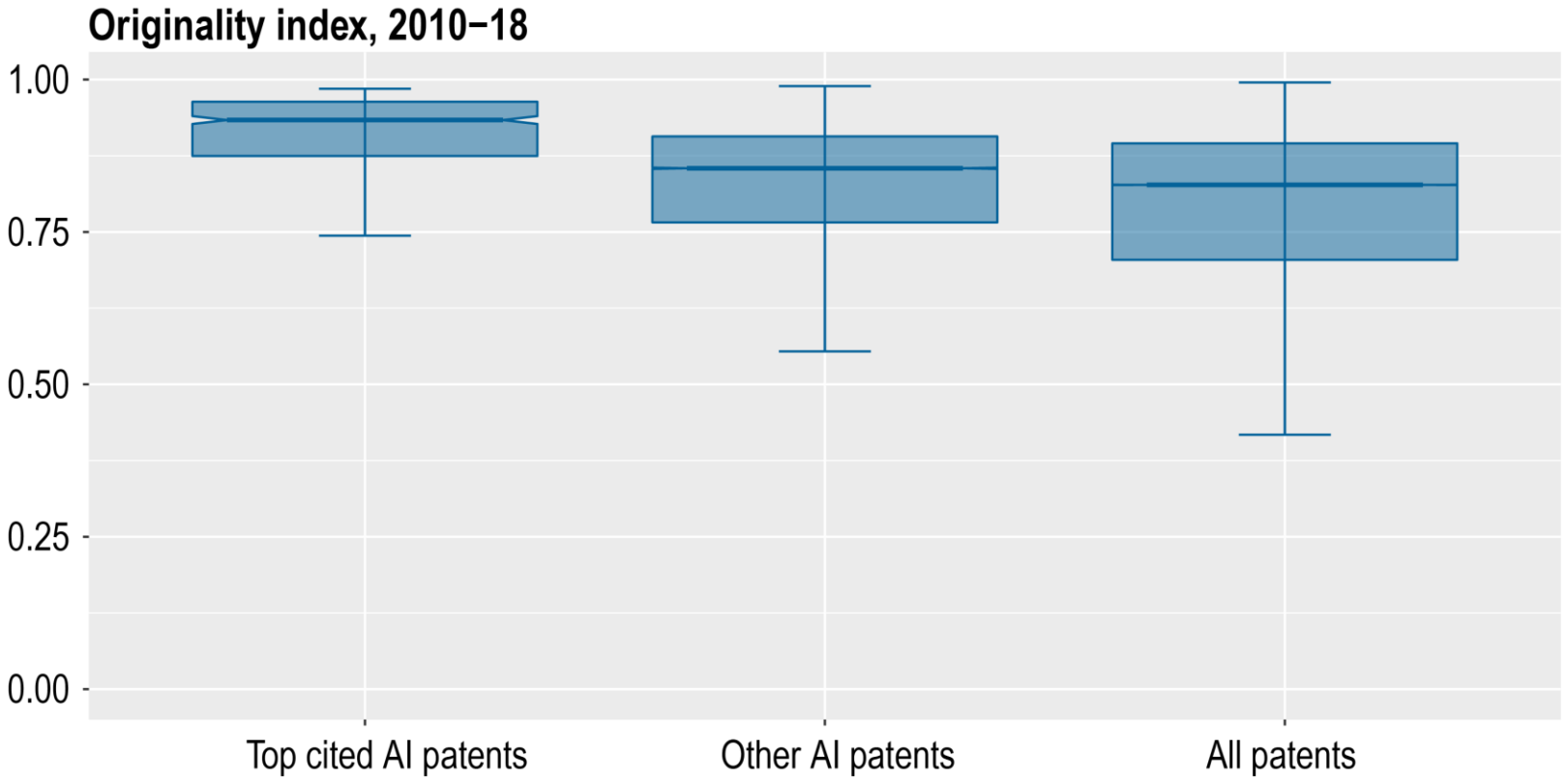


*The **patent scope index** measures the scope of a patent in terms of the number of distinct 4-digit subclasses of the IPC the invention is allocated to.*

Note: The AI patent sample (up to 553 patents) refers to the top 1% cited patent applications (normalised counts of forward citations) filed at the USPTO (patents issued and pre-grant publications) that refer to AI technologies. The patent scope index is constructed based on Squicciarini, Dernis and Criscuolo (2013).
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

… and they belong to a wide(r) range of fields

- Core AI patents also rely on a large(r) number of **diverse knowledge sources** (i.e. patents belonging to a wide(r) array of technology fields), providing evidence for AI's **GPT nature**.
- This is the case across the entire time period observed.

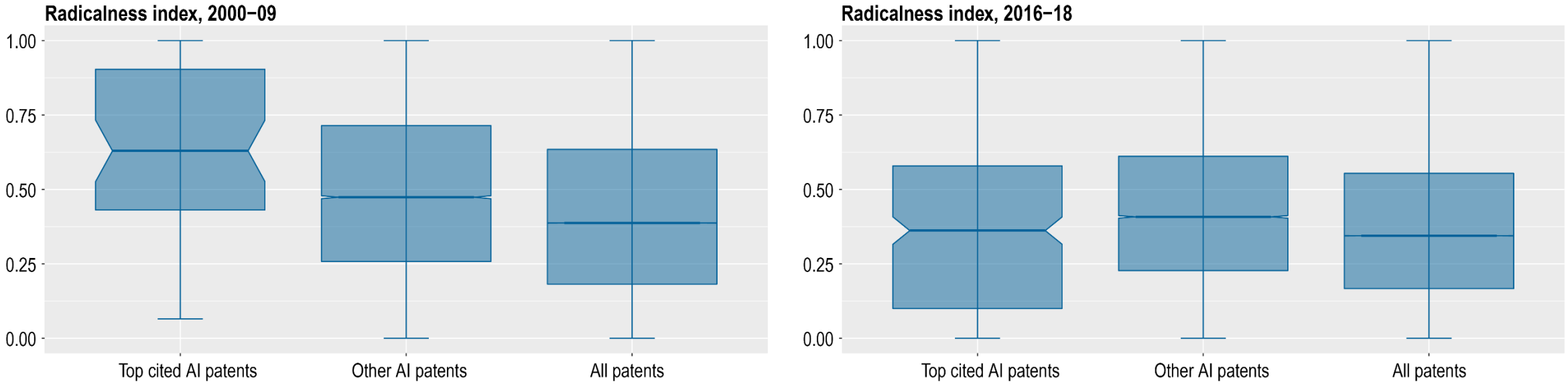


Patent originality *refers to the breadth of the technology fields on which a patent relies. Similar to the generality index measure but relying on backward rather than on forward citations.*

Note: The AI patent sample (up to 553 patents) refers to the top 1% cited patent applications (normalised counts of forward citations) filed at the USPTO (patents issued and pre-grant publications) that refer to AI technologies. The originality index is constructed based on Squicciarini, Dernis and Criscuolo (2013).
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

… but core AI patents used to contain more radical inventions

- Inventions in core AI patents **were technologically more radical** compared to overall patents, i.e. they cited previous patents in classes other than the ones they were in, and hence contained inventions that were considered more radical by building upon paradigms that differ from the one to which they are applied.
- However, this was mainly the case in earlier years and is **no longer** the case in most recent years, possibly due to AI's **GPT nature**.



*The **Radicalness index** is fundamentally backward-looking in nature as it captures the radicalness of a patent in terms of the extent to which it differs from the predecessors it relies upon.*

Note: The AI patent sample (up to 553 patents) refers to the top 1% cited patent applications (normalised counts of forward citations) filed at the USPTO (patents issued and pre-grant publications) that refer to AI technologies. The radicalness index is constructed based on Squicciarini, Dernis and Criscuolo (2013).
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, November 2022.

Australian Government
IP Australia

31 May 2023

Artificial Intelligence

IP Australia – Patent Analytics Hub

Tom Millist

P TM D PBR

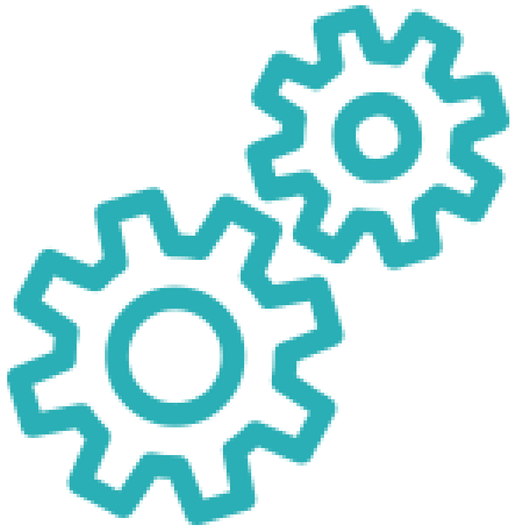
Outline

- Why?
- How?
- What?

Open discussion / Q&A

Why?

Industry



Government

Australian Government

IP Australia

Department of Industry, Science and Resources

P TM D PBR

How?

Search strategy

- Generate a search strategy to cover the area of interest

Output

- Search strategy with patent classification symbols (CPC/IPC) and keywords

Searching – EPOQUE or Derwent Innovation (DI)

- Extract patent publication numbers and publication dates

Output

- DI - CSV file
- EPOQUE – XFT file

AWS SageMaker data extraction

- Extract relevant bibliographic data from PATSTAT and IPSTAR
- Apply pre-processing steps e.g. name harmonisation

Output

- CSV file with clean data

Communication

- Tableau or PowerBI

Output

- Interactive visualisation templates

P TM D PBR

Search strategy

Patent analytics and R&D tax incentive analysis (Department of Industry, Science and Resources)

AI technologies

- Neural networks
- Learning methods
- Expert systems
- Hardware

AI applications

- Electrical engineering
- Instruments
- Chemistry
- Mechanical engineering
- Other fields

Epoque searching

Keywords and patent classifications

PATENW database

Patent publications

Topic: Expert Systems	
Subtopic	**Search statement**
Other expert systems **(ES_OT)**	G06N5/003/IC/CI/LOW NOT (OR G06N3/00, G06N20/00)/IC/CI/LOW
Inference engine **(ES_IE)**	((OR G06N5/04, G06N7/005, G06N7/06, G06N7/08)/IC/CI/LOW OR (OR G06N7/02, G06N7/023, G06N7/026)/IC/CI/LOW OR (G06N5/00/IC/CI/LOW AND (OR ((OR BACKWARD?, FORWARD?, FUZZY, ENGINE) 1W INFERENC+), ABDUCTION, (PROBABILI+ 1W REASONING), (BAYESIAN 1W (OR NETWORK+, REASONING)))/AB/DESC/CLMS/IW/TI)) NOT (OR G06N3/00, G06N20/00)/IC/CI/LOW
Knowledge Base **(ES_KB)**	(G06N5/02/IC/CI/LOW OR (G06N5/00/IC/CI/LOW AND (KNOWLEDGE 1W (OR REPRESENTATION, ENGINEERING, ACQUISITION))/AB/DESC/CLMS/IW/TI)) NOT (OR G06N3/00, G06N20/00)/IC/CI/LOW

Data extraction

Patent publications

PATSTAT

INPADOC family information

Select First Psn Name	Appln Id	Inpadoc Family Id	Day of EARLIEST_FILING_DATE	APPLN_NR_EPODOC
ZHANG XIAO	517948499	517948499	5/15/2019	CN201910410797
ZHANG XIAO	517948499	517948499	5/15/2019	CN201910410797
NÂ·æ‹‰æ ¼å¡”	527347552	496629244	6/29/2017	CN201880043730
NÂ·æ‹‰æ ¼å¡”	527347552	496629244	6/29/2017	CN201880043730
NÂ·æ‹‰æ ¼å¡”	527347552	496629244	6/29/2017	CN201880043730
NÂ·æ‹‰æ ¼å¡”	527347552	496629244	6/29/2017	CN201880043730
VERMA, ISHAN	477106897	477106897	12/12/2016	EP20170164321
VERMA, ISHAN	477106897	477106897	12/12/2016	EP20170164321
BENOIT, GILLES J.	538278684	519826004	3/15/2019	WO2019US50691
BENOIT, GILLES J.	538278684	519826004	3/15/2019	WO2019US50691
BENOIT, GILLES J.	538278684	519826004	3/15/2019	WO2019US50691
BENOIT, GILLES J.	538278684	519826004	3/15/2019	WO2019US50691
WANG, BAO HUA	478346578	478346578	11/11/2015	US201514938050
WANG, BAO HUA	478346578	478346578	11/11/2015	US201514938050
WANG, BAO HUA	478346578	478346578	11/11/2015	US201514938050
WANG, BAO HUA	478346578	478346578	11/11/2015	US201514938050

Name harmonisation

Step 0: PATSTAT name harmonisation

- Name harmonisation in database

Step 1: Basic level name harmonization

- SageMaker script to perform further changes

Step 2: Orbis name harmonization

- Search Orbis database to find the Global ultimate owner (GUO)
- Match Orbis data to PATSTAT data using Power Query

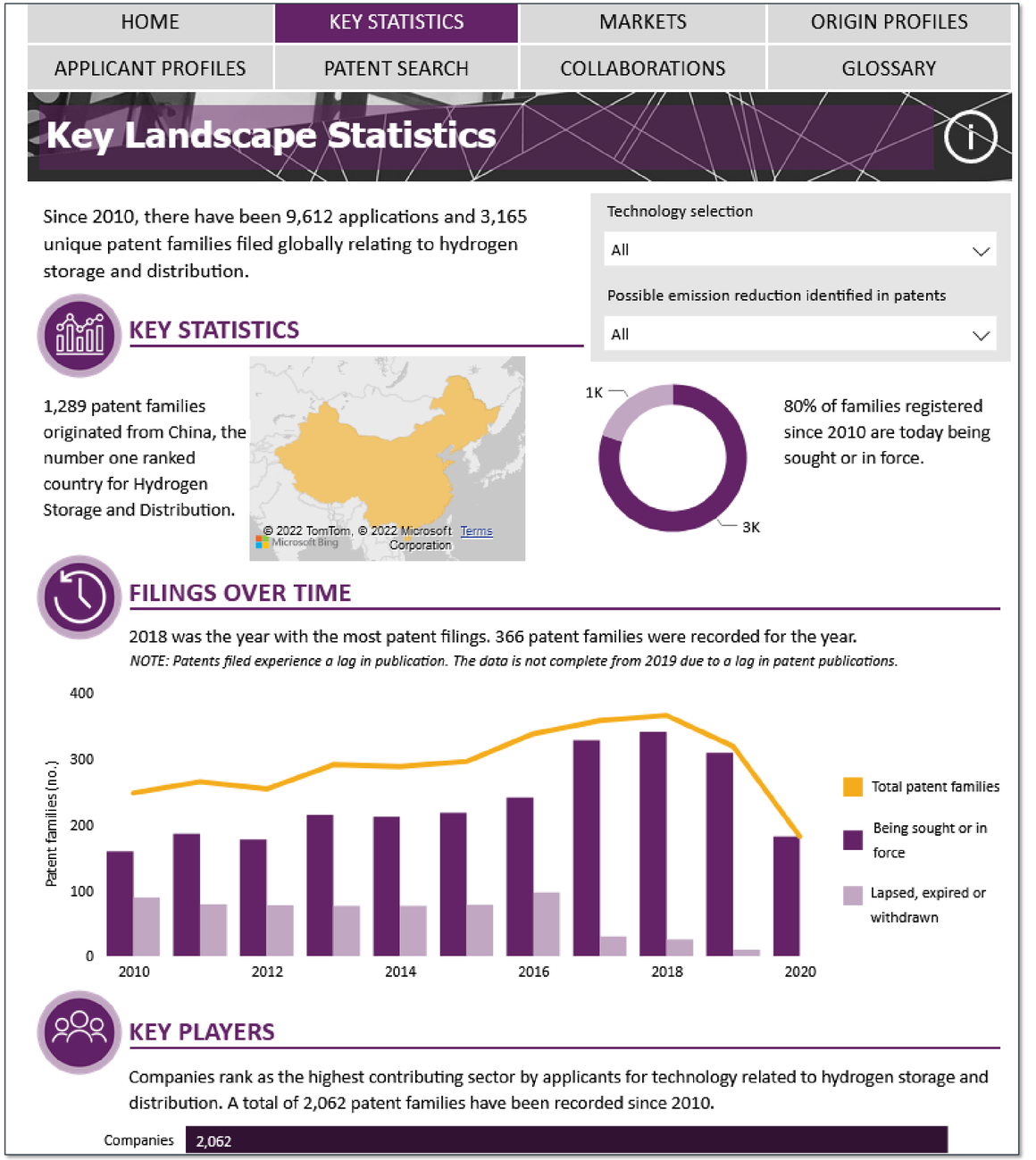
Phrase/character in the input	Replacement/remove
-	" " (space)
INC.	"" (remove)
'[,.;\'\"?!@#$%^()\[\]\{\}\/\\<\>\/_|]'	"" (remove)

orbis intellectual property

Communication

Key facts first

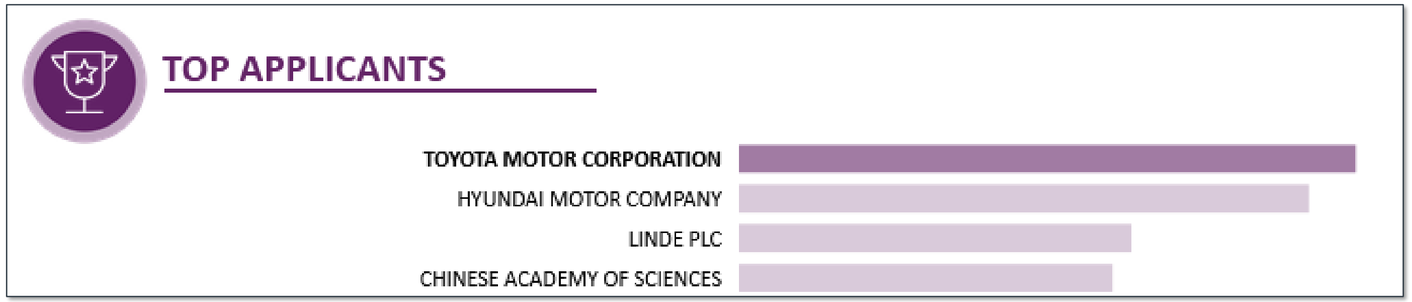
- Filings
- Peak periods
- Countries
- Business sectors/applicants of interest

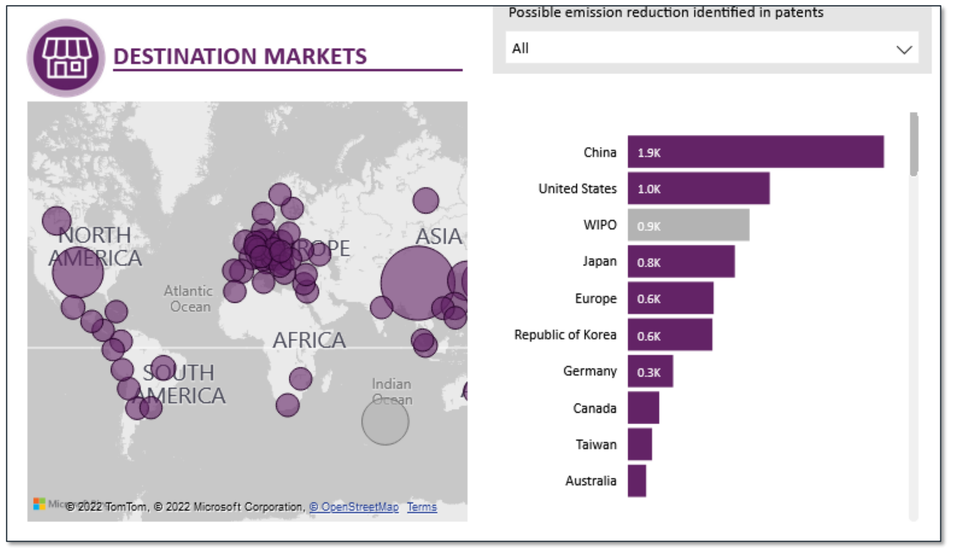


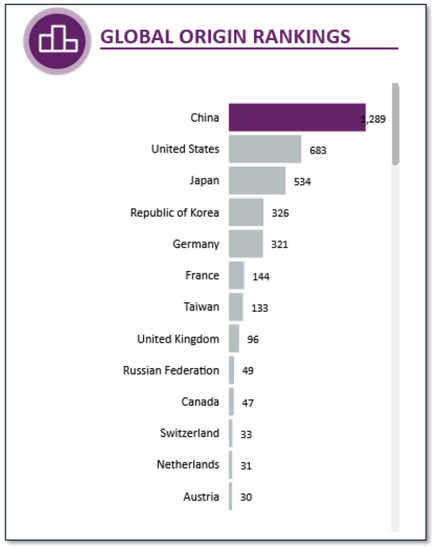
Communication

Delving in

- Markets / filing destinations
- Countries / regions of origin
- Leading applicants
- Collaboration
  - applicants
  - business sectors
  - countries
- Links to patents

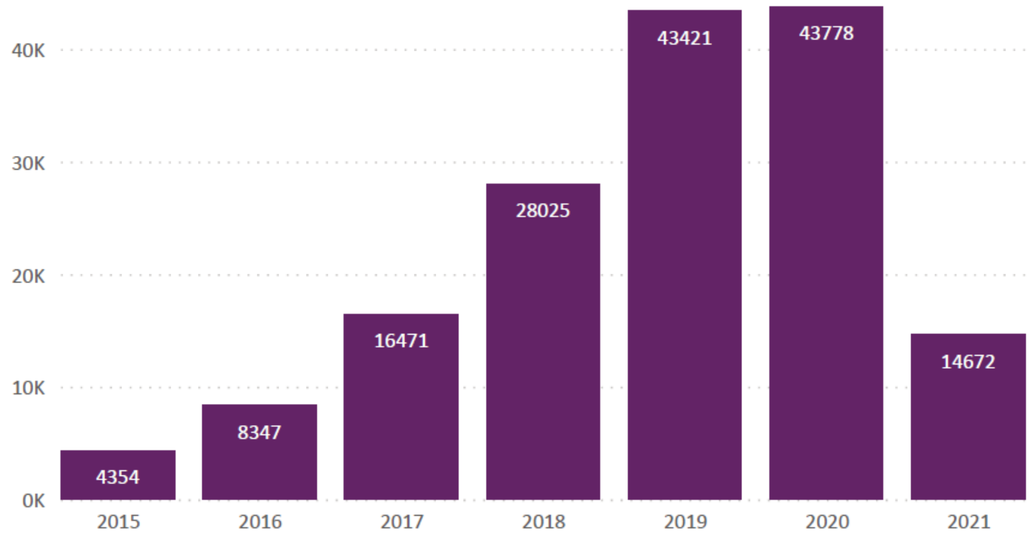


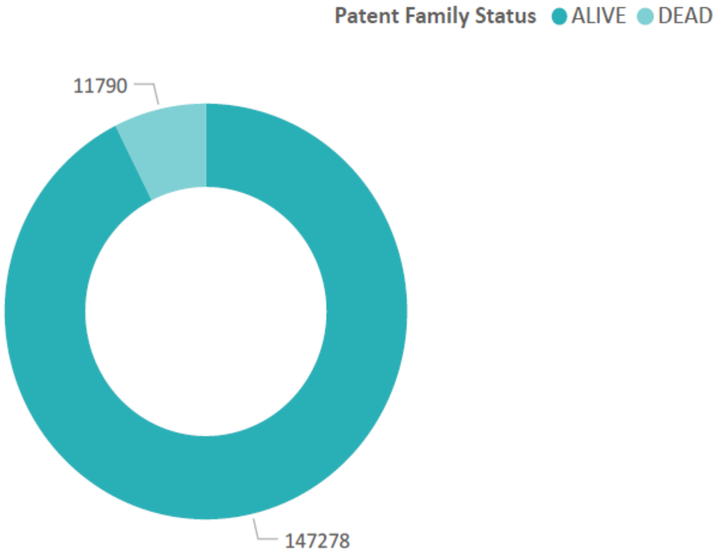




What?

AI patenting - global filings



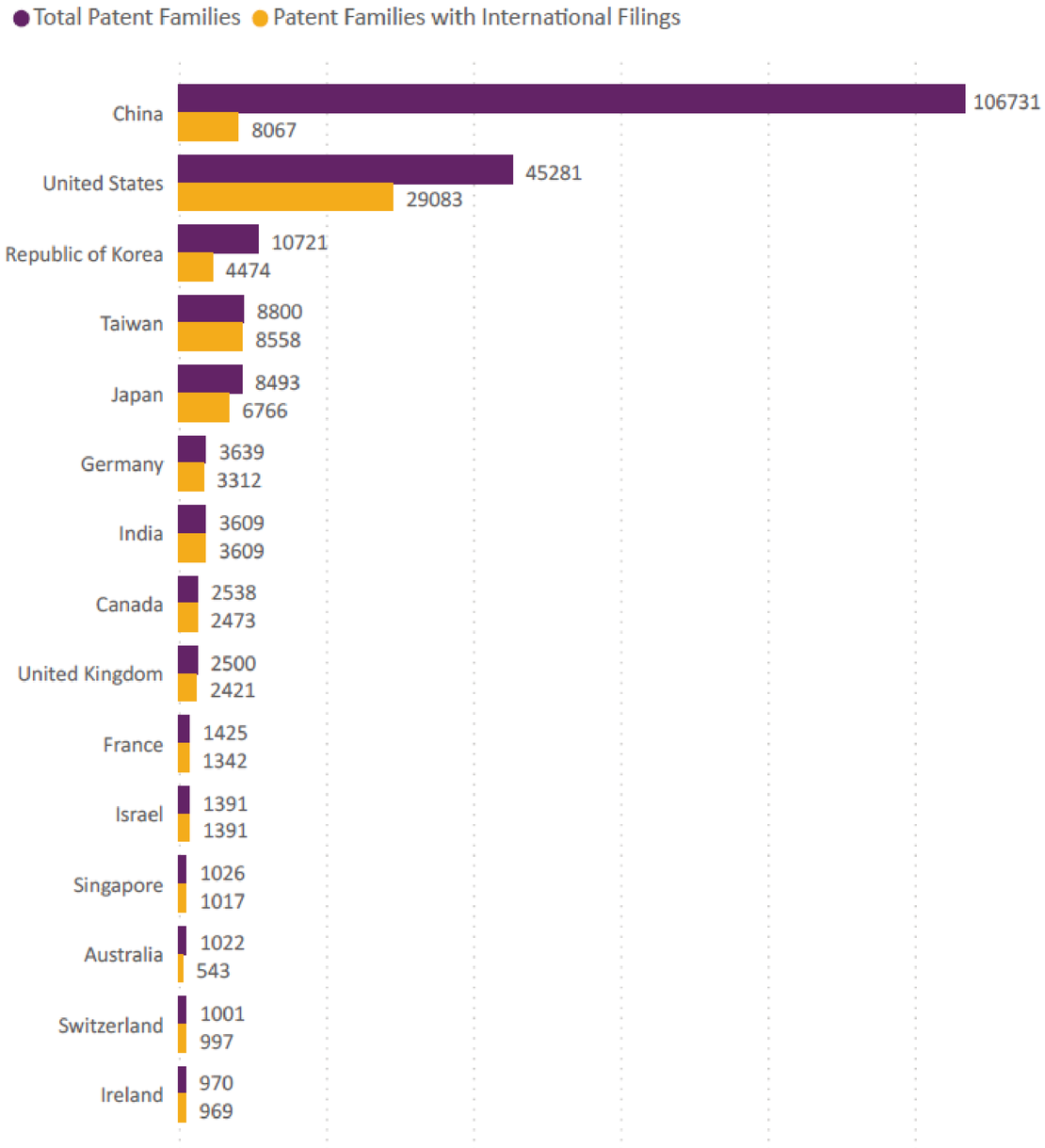


Growth: 92% (2016), 97% (2017)

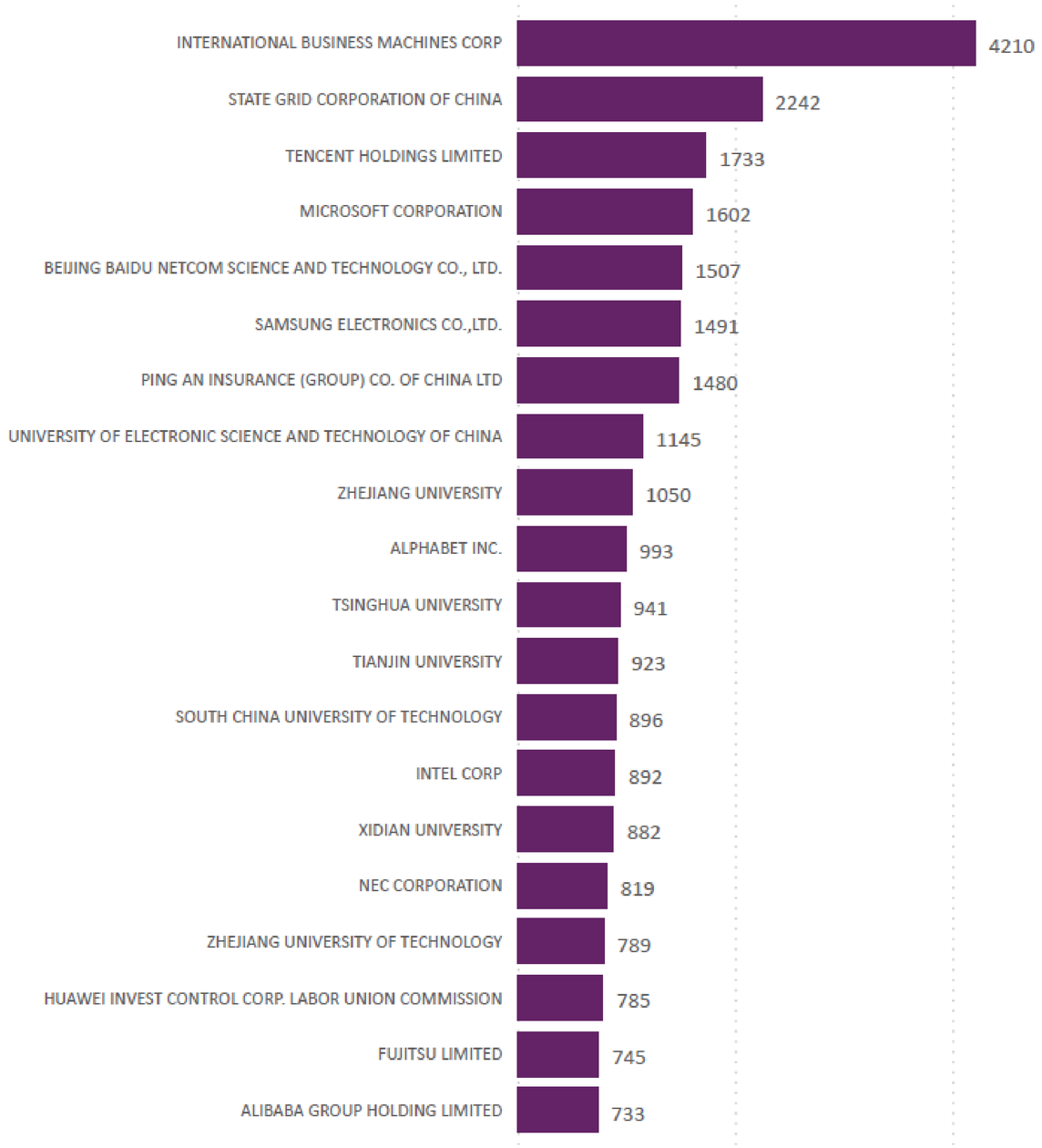
- 159,068 filed (2,176 in AU)
- 133,562 granted
- 93% alive

P TM D PBR

Origin


Total Patent Families
Patent Families with International Filings
China
106731
8067
United States
45281
29083
Republic of Korea
10721
4474
Taiwan
8800
8558
Japan
8493
6766
Germany
3639
3312
India
3609
3609
Canada
2538
2473
United Kingdom
2500
2421
France
1425
1342
Israel
1391
1391
Singapore
1026
1017
Australia
1022
543
Switzerland
1001
997
Ireland
970
969

Applicants



P TM D PBR

Where are Australians filing?

Filing Jurisdiction	Patent Families
United States	333
Australia	179
China	97
Europe	68
Canada	28
Republic of Korea	24
Japan	21
Singapore	14
Taiwan	10
United Kingdom	8
Germany	7
Chile	4
Brazil	3
Israel	3
Philippines	3
Denmark	1
Russian Federation	1
Sweden	1

P TM D PBR

Australian Government
IP Australia

Thank you

Discussion / Q&A

analytics@ipaustralia.gov.au

P TM D PBR