Panel "Developing and applying AI: Core and non-core AI", Tuesday 2nd February 2021

#OECDAI

The 5 Ws of Artificial Intelligence:

Identifying AI Agents and Characterising AI Developments

Mariagrazia Squicciarini *

OECD Directorate for Science, Technology and Innovation

* Based on work with: Emile Cammeraat, Hélène Dernis, Laurent Moussiegt, Heike Nachtigall, Daisuke Nawa and Lea Samek

International Conference on AI in Work, Innovation, Productivity and Skills

1-5 February 2021

OECD

Bundesministerium für Arbeit und Soziales

UNDERSTANDING AI GENERATION AND DIFFUSION: THE 5 W'S OF AI

Designing and implementing effective AI policies needs relying on evidence about :

- **WHO** >>> AI agents (firms, universities, PROs, etc.) creating and/or adopting AI-related technologies.
- **WHAT** >>> "Core" AI developments versus "Non-Core" AI advances, i.e. mainly applications, and their relationship with economic performance and societal wellbeing.
- **WHERE** >>> AI actors (including leaders) worldwide, i.e. both countries and agents, their location, organisational setting, etc..
- **WHEN** >>> evolution over time, to identify progresses, (co)evolution and divergences, etc..
- **WHY** >>> what enables or hinders AI developments, i.e. technological specialisation, human capital and broader framework conditions needed for people centered AI.

IDENTIFYING AI ACTORS USING PATENT AND TM DATA

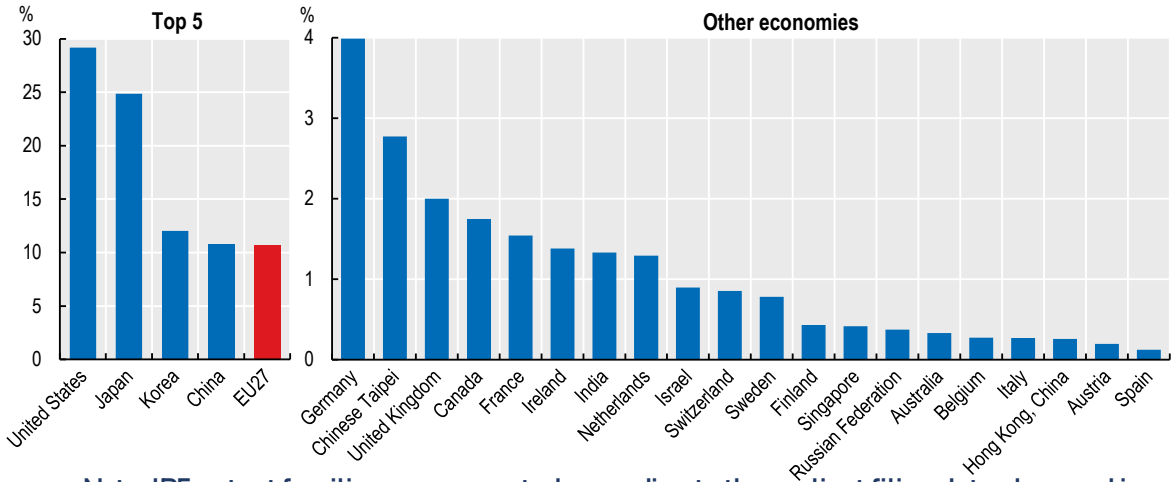
COUNTRIES

Use info based on:

- AI-related patents (Baruffaldi et al., 2020), using combinations of keywords and International Patent Classification (IPC) and Cooperative Patent Classification (CPC) classes;
- AI-related trademarks (Nakazato and Squicciarini, forthcoming), using AI-related keywords information contained in reports from JPO, the UK IPO, WIPO, as well as Baruffaldi et al. (2020).

TOP 25 ECONOMIES WITH AI PATENTS, 2014-18

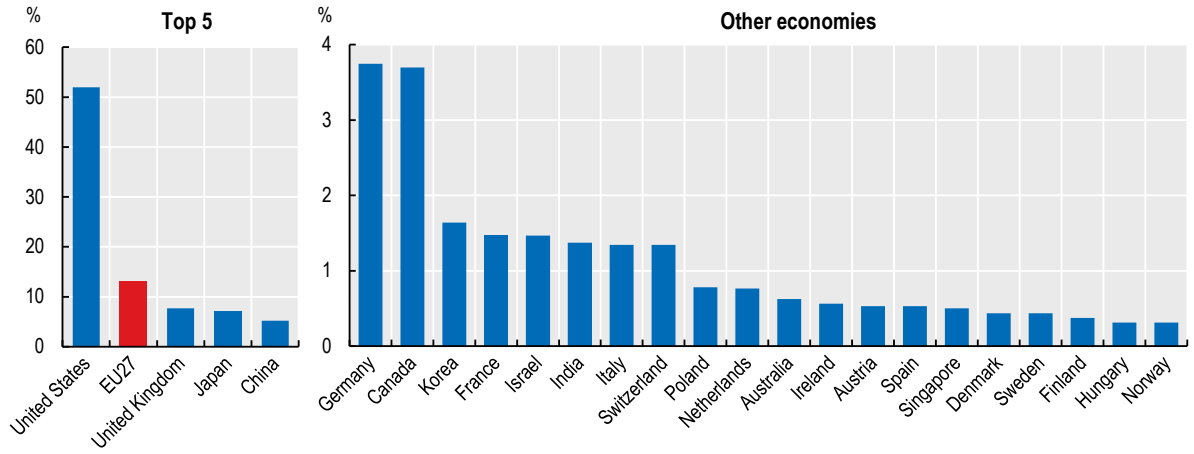
Economies' share in total IP5 patents families in AI



Note: IP5 patent families are presented according to the earliest filing date observed in the family and the location of applicants listed in the family, using fractional counts.

TOP 25 ECONOMIES WITH AI TRADEMARKS, 2014-18

Economies' share in total AI-related TM filed at EUIPO, JPO and USPTO



Note: Trademarks data refer to registrations at the EUIPO, JPO and USPTO, by filing date and location of the applicants, using fractional counts.

Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, January 2021

IDENTIFYING AI ACTORS USING PATENT AND TM DATA

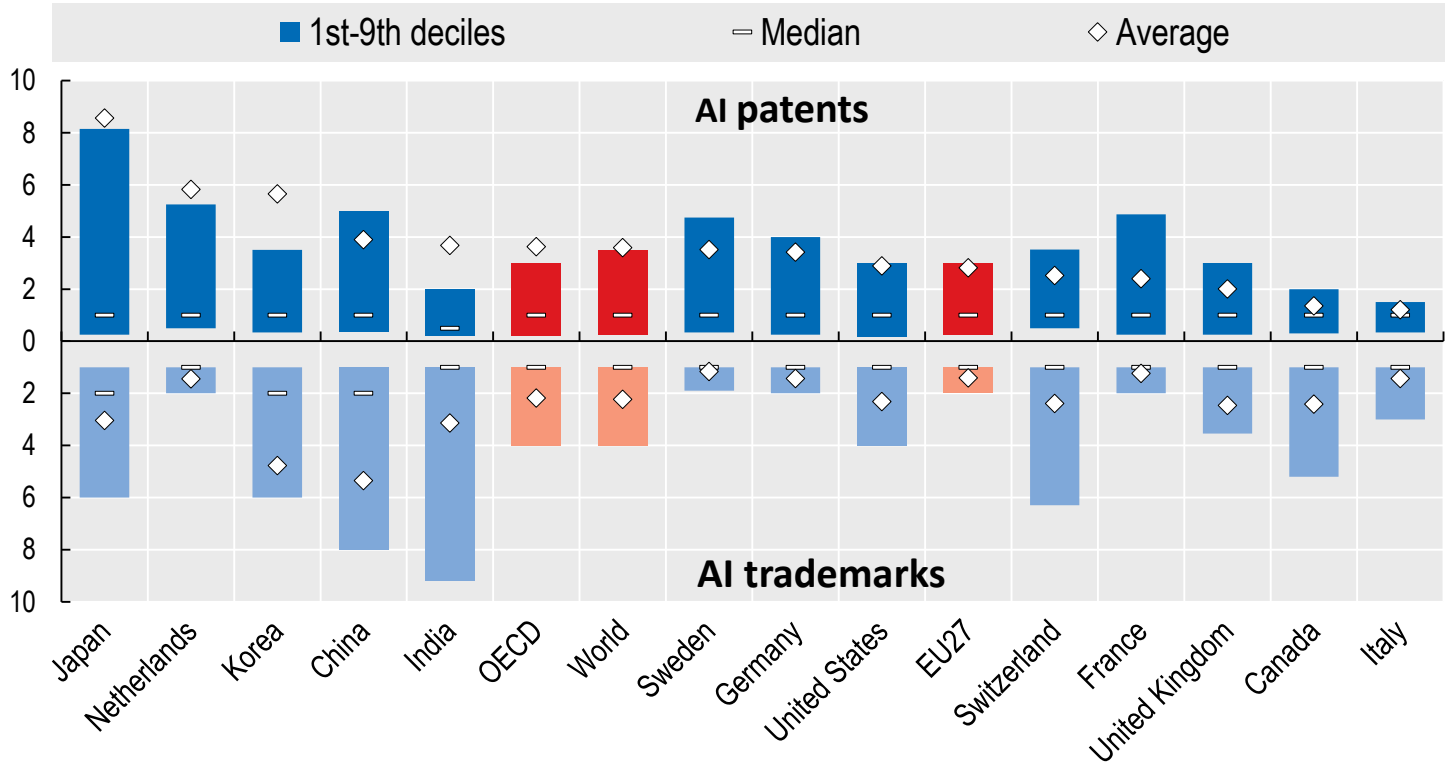
COMPANIES

TOP 50 APPLICANTS WITH IP ASSETS IN AI

Share of companies' IP portfolio in AI in total patents and TMs in AI

AI-RELATED PATENTS AND TM, BY APPLICANTS, 2014-18

Average number of patents and TMs per applicant, median and deciles distribution



Note: Only economies with 40+ applicants with AI patents and 10 applicants with AI TMs are included

Source: Dernis, Moussiegt, Nawa and Squicciarini.(forthcoming). OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, January 2021

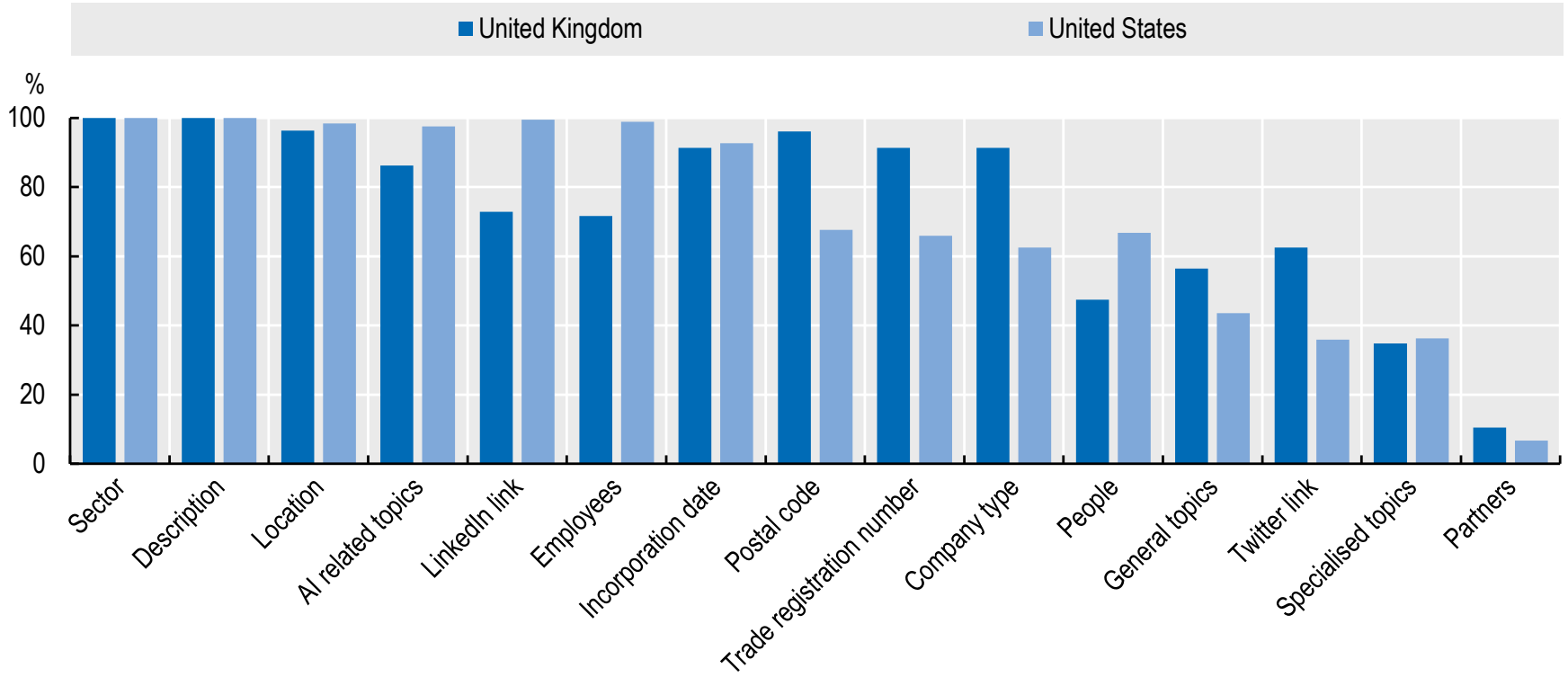
Identifying AI Actors Reading the Web at Scale

Based on experimental data collected by GlassAI for the OECD. GlassAI reads and interprets open web text at scale, exploiting any machine-readable information (e.g. sentences, paragraphs, etc.) contained on websites.

AI agents (companies, universities, etc.) identified using keywords from Baruffaldi et al. (2020).

Share of Companies Featuring Relevant Pieces of Information, UK & US

By main variable



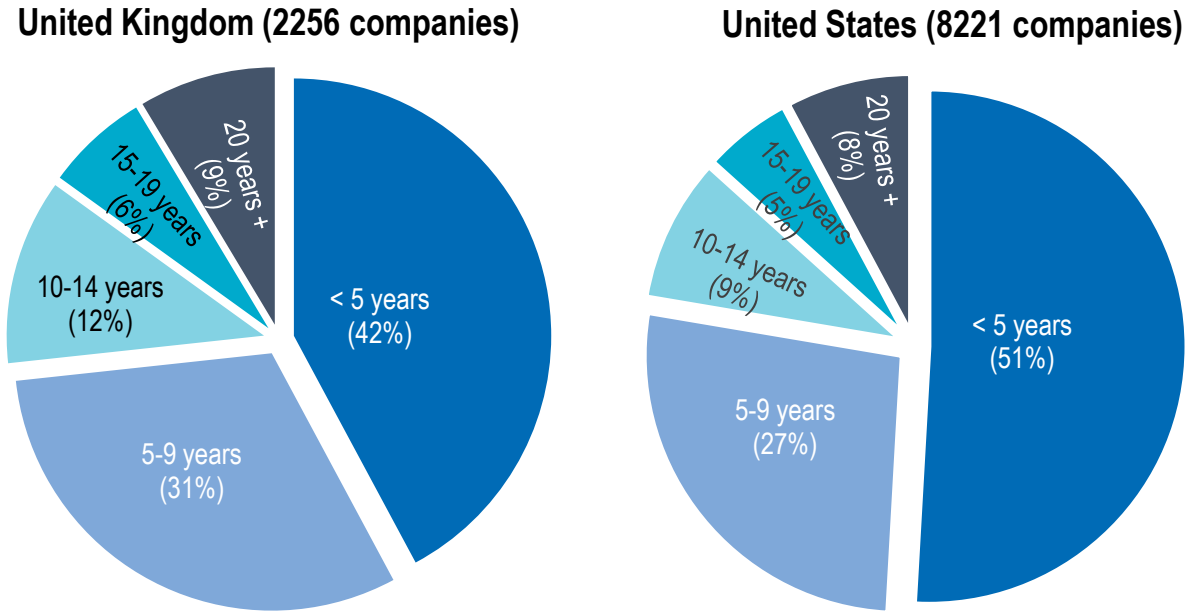
Source: Dernis, Moussiegt, Nawa and Squicciarini (forthcoming). OECD calculations based on GlassAI data, January 2021

Characterising AI Actors Reading the Web at Scale

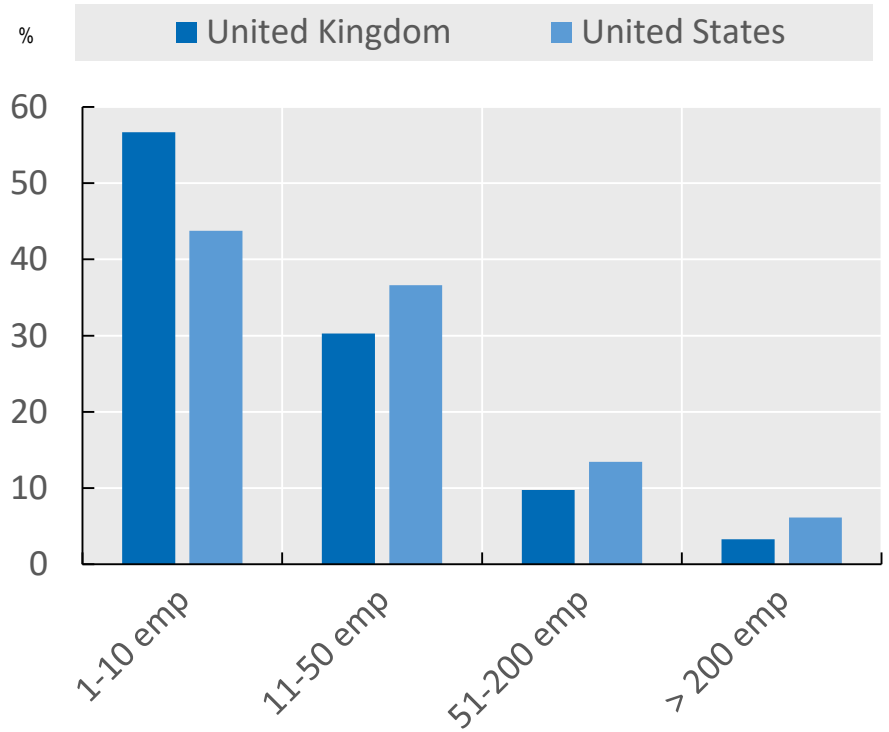
Based on experimental data collected by GlassAI for the OECD. GlassAI reads and interprets open web text at scale, exploiting any machine-readable information (e.g. sentences, paragraphs, etc.) contained on websites.

AI agents (companies, universities, etc.) identified using keywords from Baruffaldi et al. (2020).

Age of AI Companies, UK & US



Size of AI Companies, UK & US

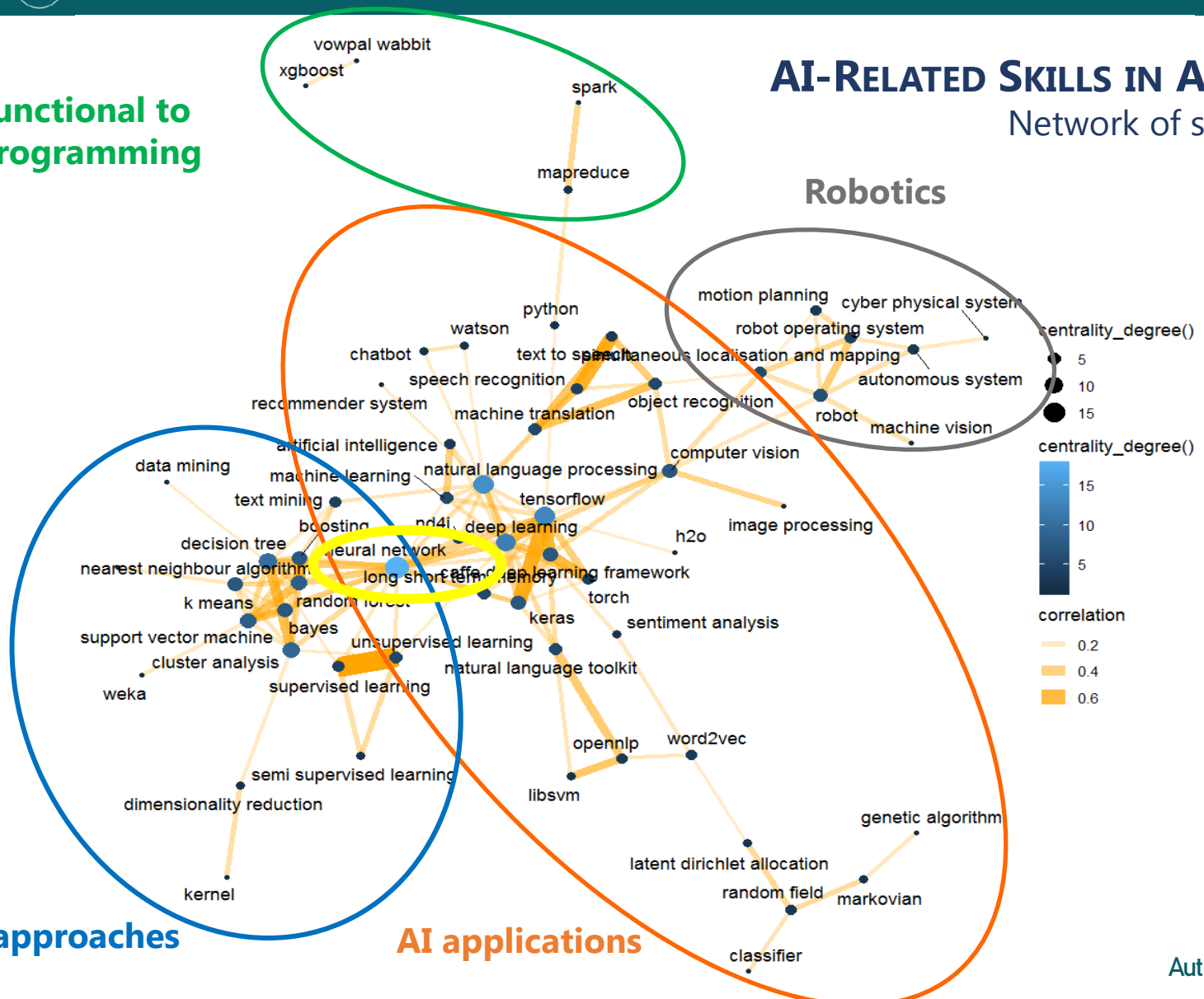


Source: Dernis, Moussiegt, Nawa and Squicciarini (forthcoming). OECD calculations based on GlassAI data, January 2021

WHAT IS "CORE" AND WHAT IS "NON-CORE"?

AI-RELATED SKILLS IN AI-RELATED JOBS, BGT DATA (US)

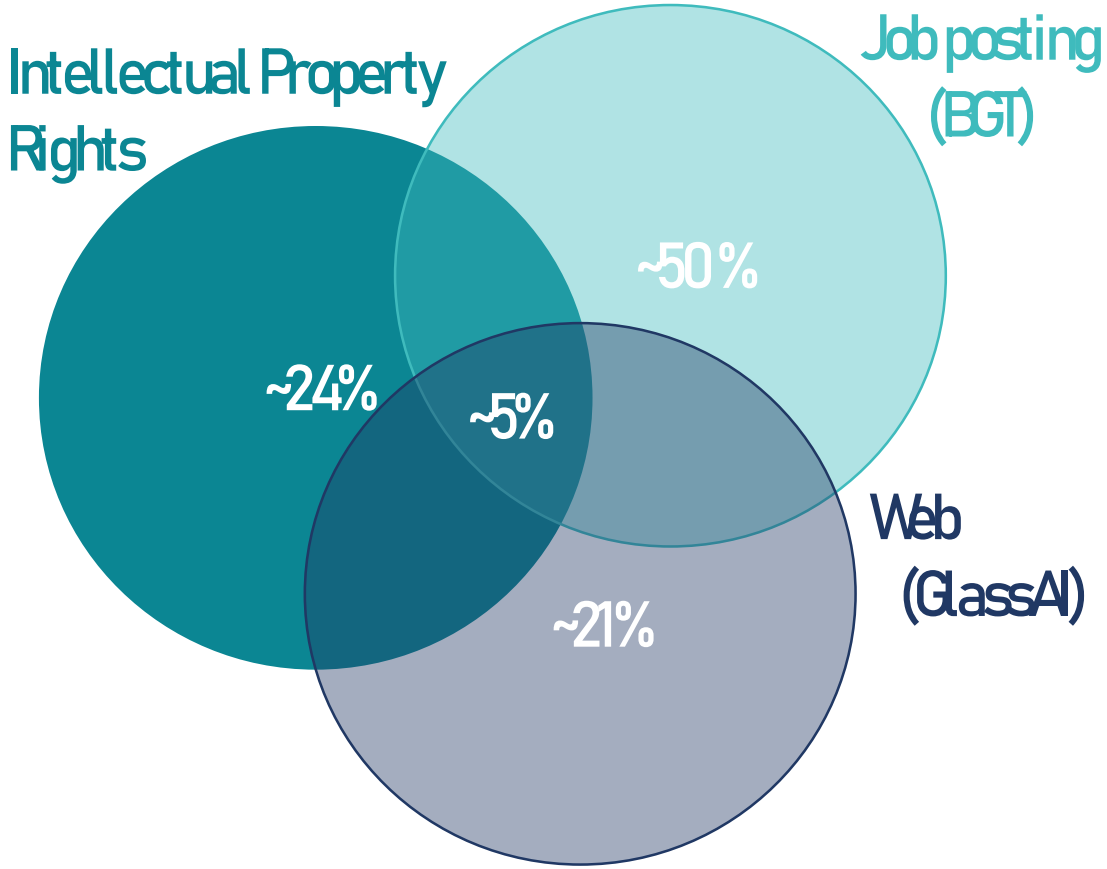
Network of skills, Corr>0.1, 2017-19



- Skill bundles are interrelated;
- Central node: Neural networks;
- A methodological / core developments part of the network, advancing AI itself, emerges (blue);
- An application-related part of the cluster emerges (orange);
- A robotics-related part of the cluster emerges (grey);
- Support functions / software bundles of skills emerge (green)

Source: Samek , Squicciarini and Cammeraat (forthcoming).
Authors' own compilation based on Burning Glass Technologies data, 2020.

MULTIPLE APPROACHES FOR A COMPLEX PROBLEM



Note: Based on data for the US and the UK only, for the period 2014-2018.

Source: OECD compilation based on BGT data, Glass AI data and STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats.

All the approaches manage to identify additional AI agents and add pieces to the AI puzzle:

- Patents protect inventions ad help identifying companies and other entities developing new AI-related technologies;
- TM allow identifying agents putting new AI-related goods and services onto the market;
- GlassAI data encompass companies and other types of institutions / organisations stating to do AI /work with AI on their websites;
- BGT data identify companies looking for AI-related human capital, for different purposes (both development and adoption).

QUESTIONS FOR DISCUSSION

None of the approaches proposed gets the full picture – and many other approaches can be pursued – but they all add important pieces of info:

1) **Who are AI agents and how to identify them?**

And, as AI development and adoption-related policies need to rely on a better understanding of what core AI developments are, as compared to non-core AI advances (i.e. mainly applications)

2) **Is a distinction between "core" and "non-core" developments really possible? And how to best distinguish core from non-core AI?**

Many thanks, mariagrazia.squicciarini@oecd.org