



EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Increasing quality of web-based data: Human's Role in production of Consistent Labour Market intelligence



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat 

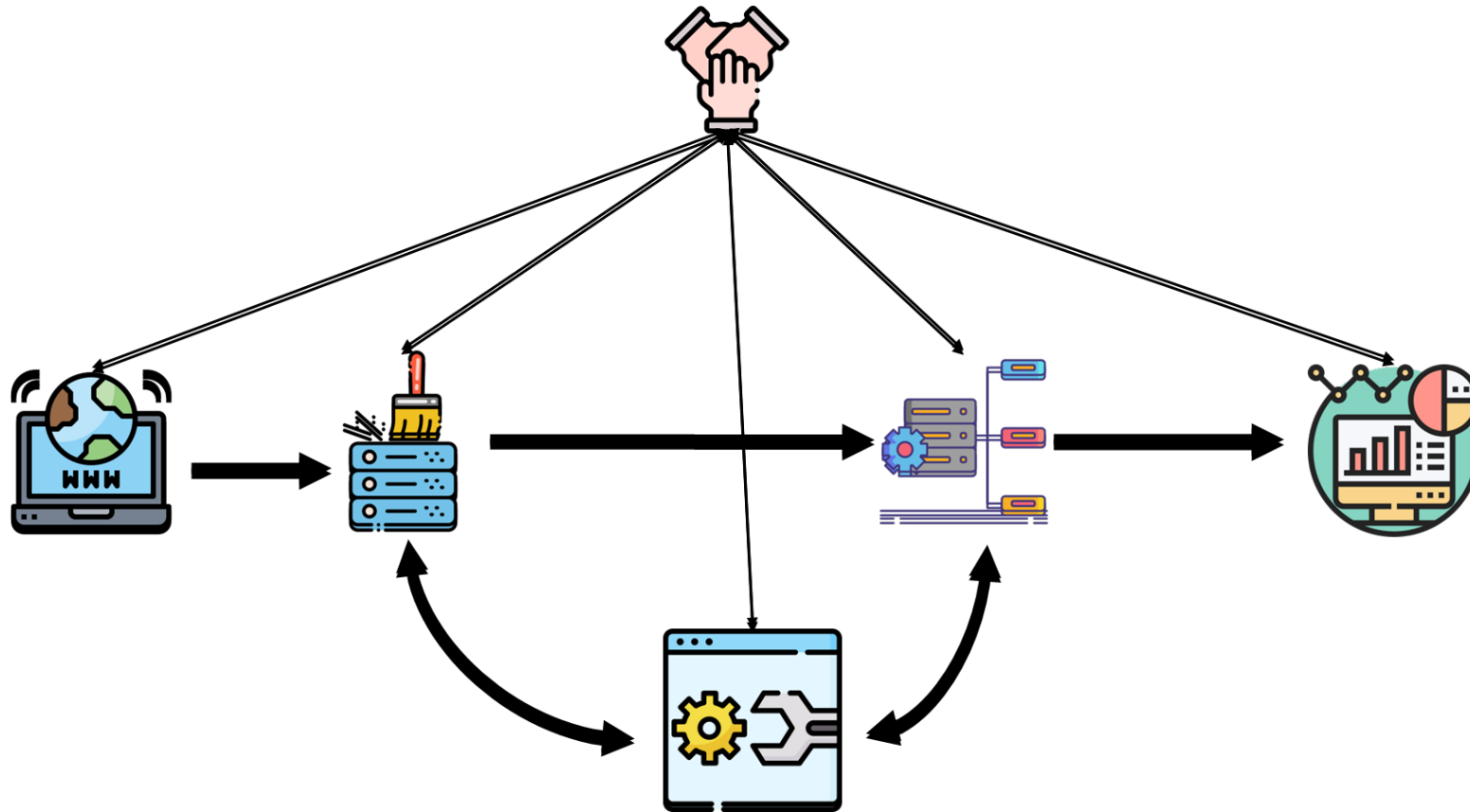
The conference is partly
financed by the European Union

Vladimir Kvetan, Jiri Branka

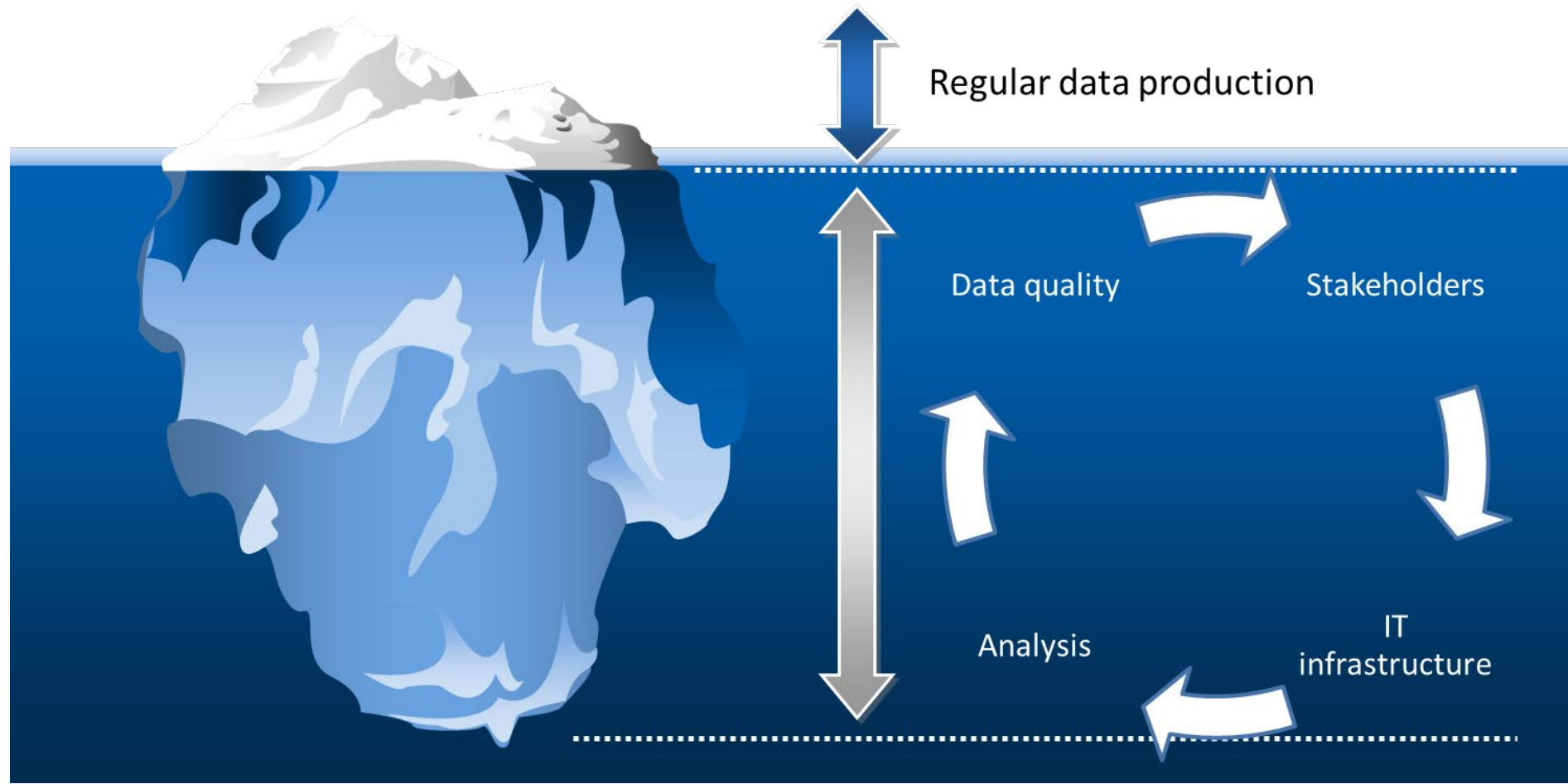
Outline of the presentation

- Introduction to WIH OJA data production and skills extractions
- Securing even skills extraction
 - Human powered
 - AI powered
- Conclusions and steps forward

Data production system secured by WIH ...

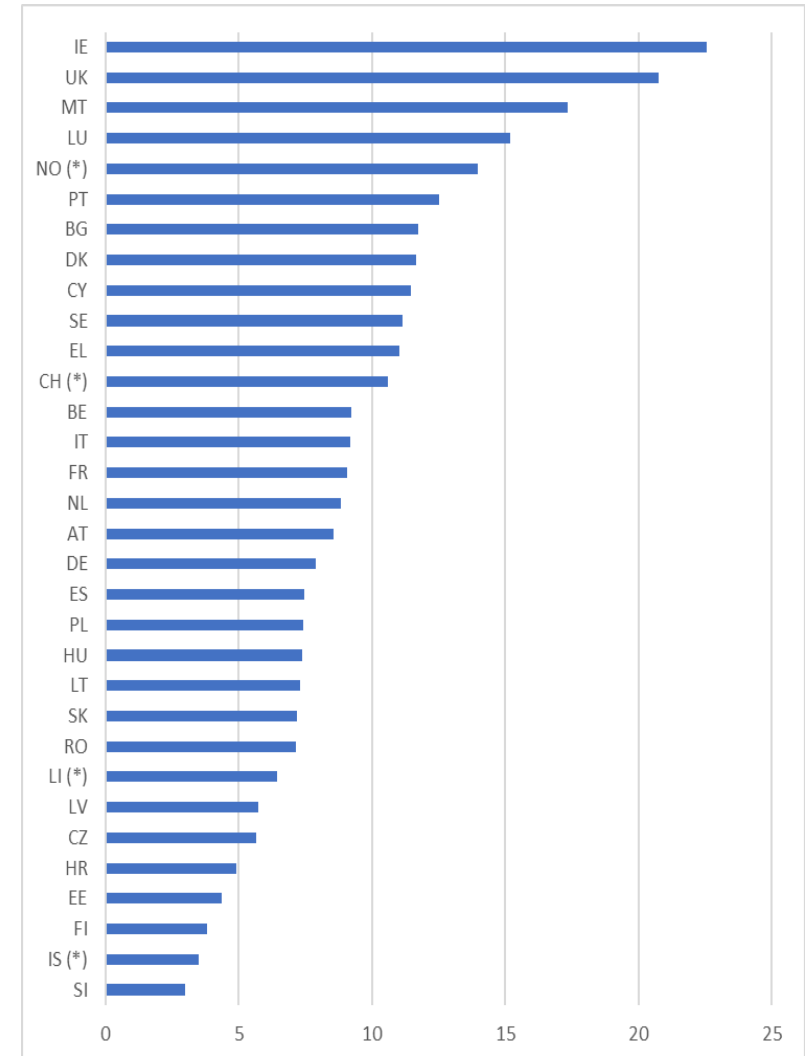


... the DPS is being constantly improved.

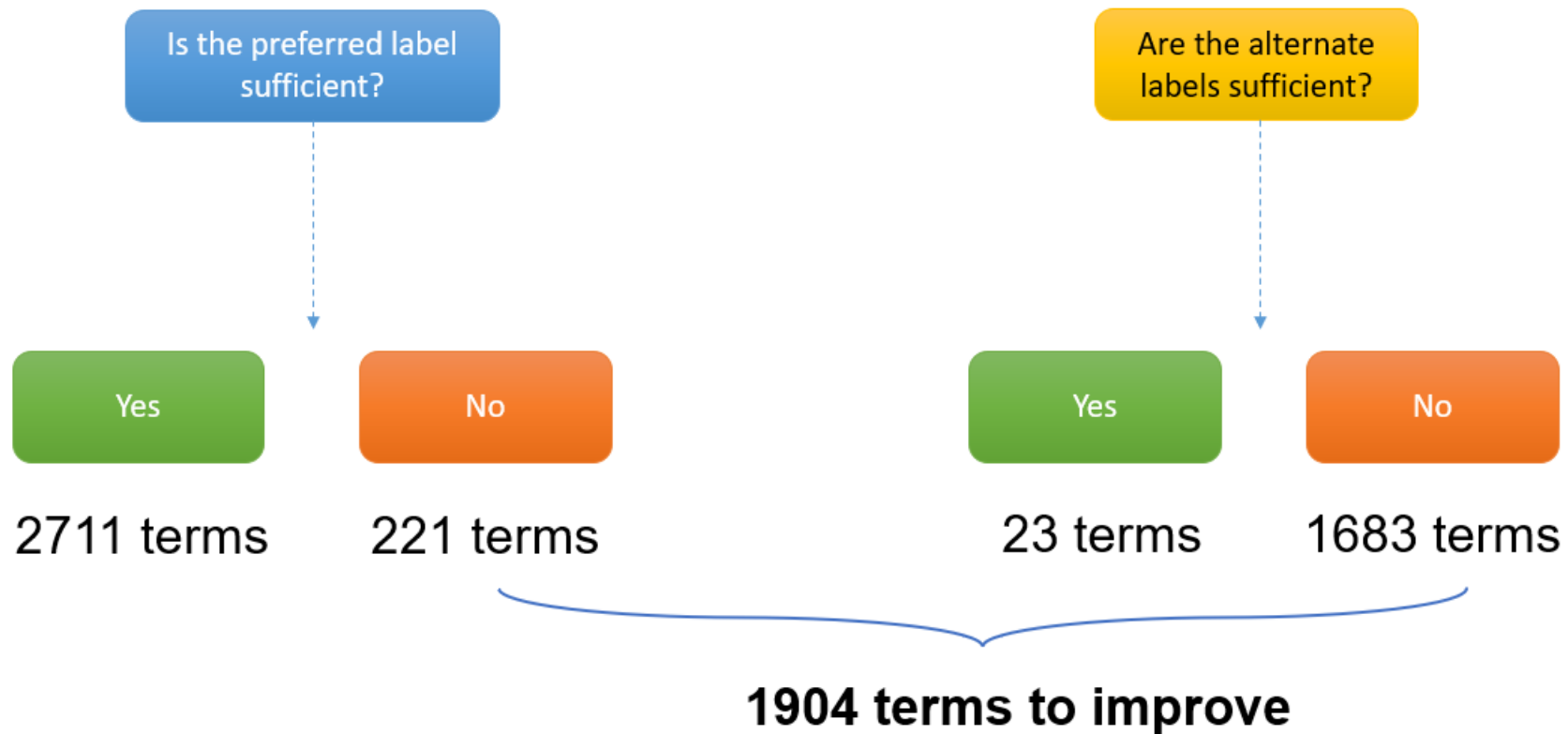


Same system, different skills extraction...

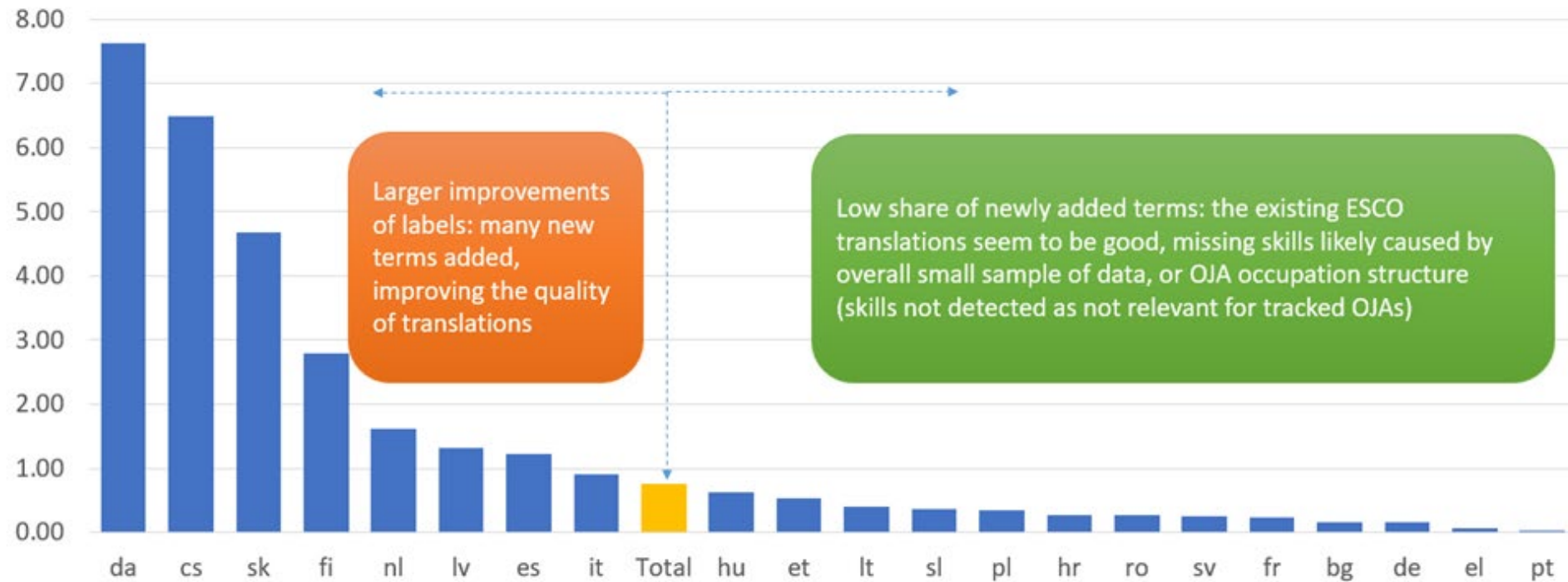
- Key characteristics
 - OJAs processed in original language
 - Ontology based extraction
 - ESCO
- The mystery of missing skills...
 - ontology?
 - algorithm?
 - OJAs?
- Two exercises focussing on ontology
 - Human based – to unify skills terms
 - AI powered – to identify “new” skills



Revision of skills ontology revealed uneven distribution of alternative skills terms...



Adding skills terms brings improvements to some language pipelines...



“New terms” identified using AI assessed by HI

Webmaster

City of Tamarac ★★★★★ 20 reviews

City of Tamarac, FL

\$59,396 - \$79,262 a year - Full-time

You must create an Indeed account before continuing to the company website to apply

Apply on company site

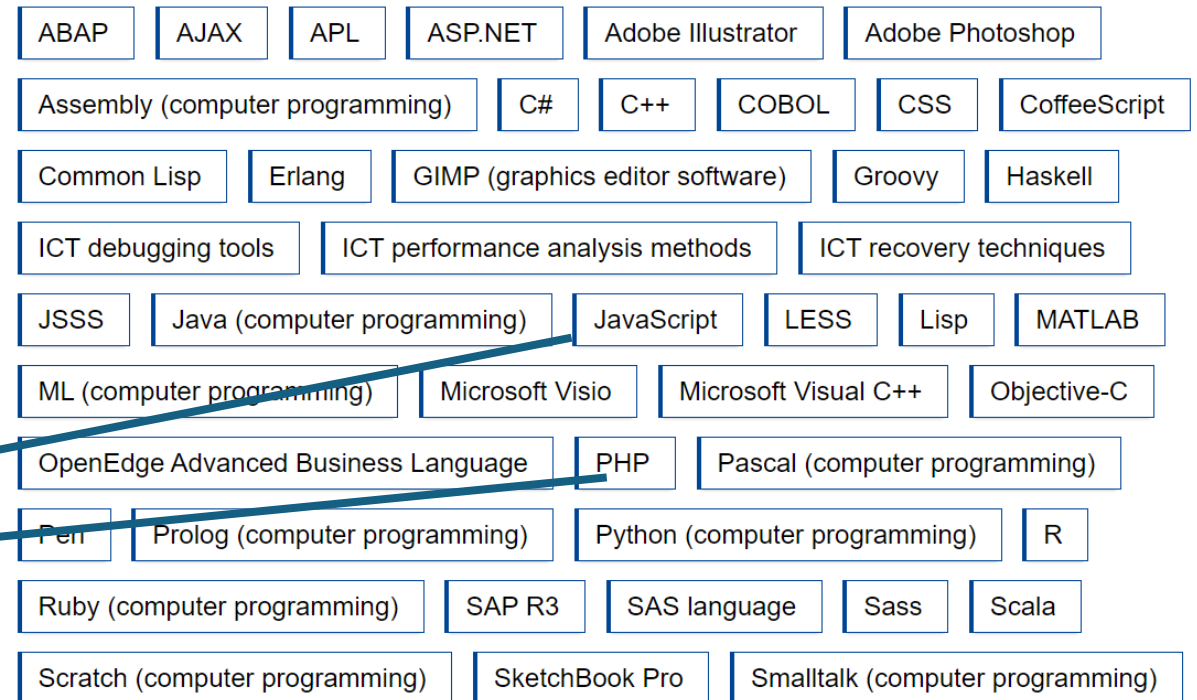


Minimum Training and Experience

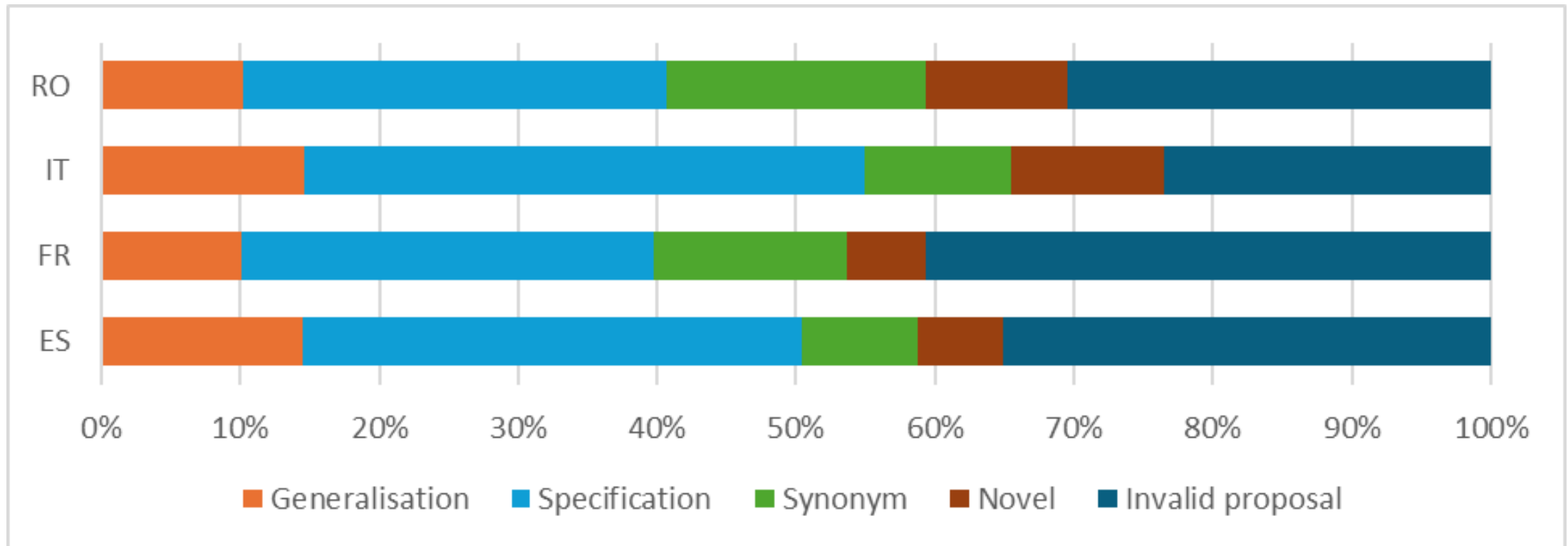
Requires a Bachelor's degree from an accredited college or university with major course work in Computer Science, Information Technology, Management Information Systems, or related field. Requires two (2) years of experience working in professional web development, including experience working with HTML, ASP, VBScript, Visual Studio.NET, CSS, PHP, JavaScript, MySQL and general network communications, Microsoft SharePoint or Microsoft SharePoint Online, Microsoft Windows Internet Information Services (IIS), Apache Tomcat and SSL Certificates, or equivalent combination of training and experience that provides the required knowledge, skills and abilities. Experience with relational databases such as Microsoft SQL Server databases is preferred.

*

ESCO list of skills for webmaster



Not all “new terms” are “new skills”





Conclusions and the next steps

- The cooperation between Eurostat and Cedefop (WIH) is bringing first results.
- There is a way to improve the quality of the data and skills extraction.
- Improvement can be powered by AI, although HI is crucial in this process.
- Next steps
 - Establish cooperation with ESCO team and establish channels for information exchange;
 - Develop “New skills” modules for all language pipelines powered by AI and LLM;
 - Seek for a ways to establish an HI expert group to feed the process.