Perspectives of using Artificial Intelligence in ESCO

34th ESCO MAI Meeting

22 October 2020

Mission of Artificial Intelligence for ESCO

Use an analytical approach based on statistical analysis, data science and machine learning to **assist** in:

1) Making maintenance of ESCO more efficient
2) Expanding ESCO
3) Making ESCO easier to use by implementers

Artificial Intelligence Use Cases for ESCO

Artificial Intelligence can support:

- ESCO skill and occupation identification from free text (e.g. work history, job descriptions)
- Linking learning outcomes to skills and occupations
- Finding close occupations and skills
- Detecting missing skills and occupations
- Detect ESCO quality issues

Artificial Intelligence Building Blocks

1. Develop algorithms to map raw, real-world data to ESCO

- ESCO content update: raw text from blueprint projects
- Learning Outcome Linking: raw text from learning opportunities to ESCO skills
- Mapping job titles to ESCO occupations: 'C++ programmer' => 'software developer'
- Identify labour market trends based on job vacancies: e.g. new jobs in a domain, …

2. Develop algorithms to quantify relations between concepts in ESCO

- Quantify relation between skills: e.g. how correlated are skills? Do you need one to have the other? Skill recommendation, …
- Quantify relation between occupation and skill
- Quantify relation between occupations
- Quantify relation between course and occupation
- Quantify relation between course and skill
- Hierarchy building and completion

3. Develop algorithms to quantify quality issues

- Missing concepts: e.g. COVID-19, IT, green
- Suspicious/missing relations between concepts
- Duplicate concepts
- Vague descriptions
- Translation issues

4. Develop algorithms for skills intelligence

- Identify skill and occupation trends
- Emerging skills and occupations

5. Develop algorithms to visualise ESCO

Artificial Intelligence Building Blocks

1. Develop algorithms to map raw, real-world data to ESCO

- ESCO content update: raw text from blueprint projects
- Learning Outcome Linking: raw text from learning opportunities to ESCO skills
- Mapping job titles to ESCO occupations: 'C++ programmer' => 'software developer'
- Identify labour market trends based on job vacancies: e.g. new jobs in a domain, …

2. Develop algorithms to quantify relations between concepts in ESCO

- Quantify relation between skills: e.g. how correlated are skills? Do you need one to have the other? Skill recommendation, …
- Quantify relation between occupation and skill
- Quantify relation between occupations
- Quantify relation between course and occupation
- Quantify relation between course and skill
- Hierarchy building and completion

3. Develop algorithms to quantify quality issues

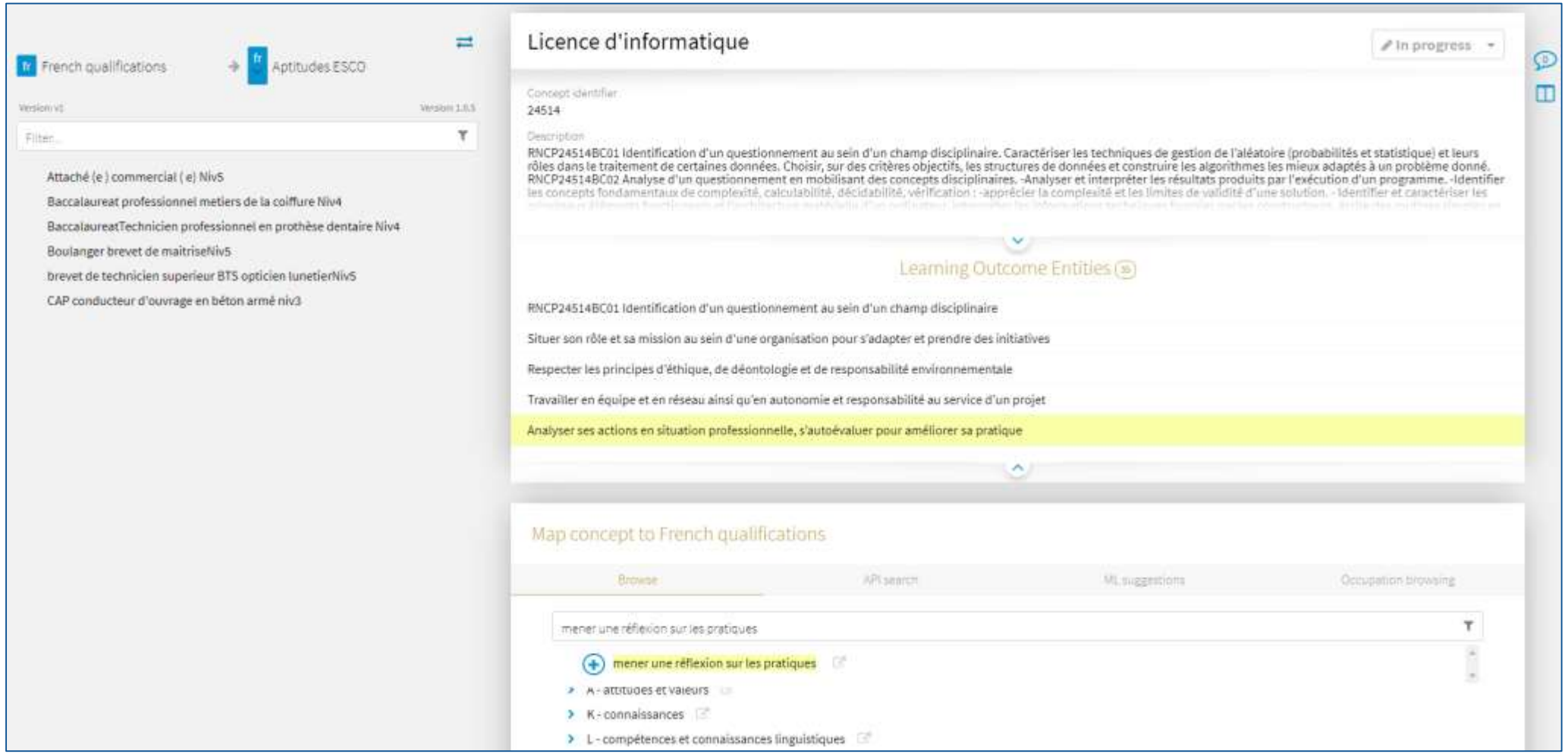
- Missing concepts: e.g. COVID-19, IT, green
- Suspicious/missing relations between concepts
- Duplicate concepts
- Vague descriptions
- Translation issues

4. Develop algorithms for skills intelligence

- Identify skill and occupation trends
- Emerging skills and occupations

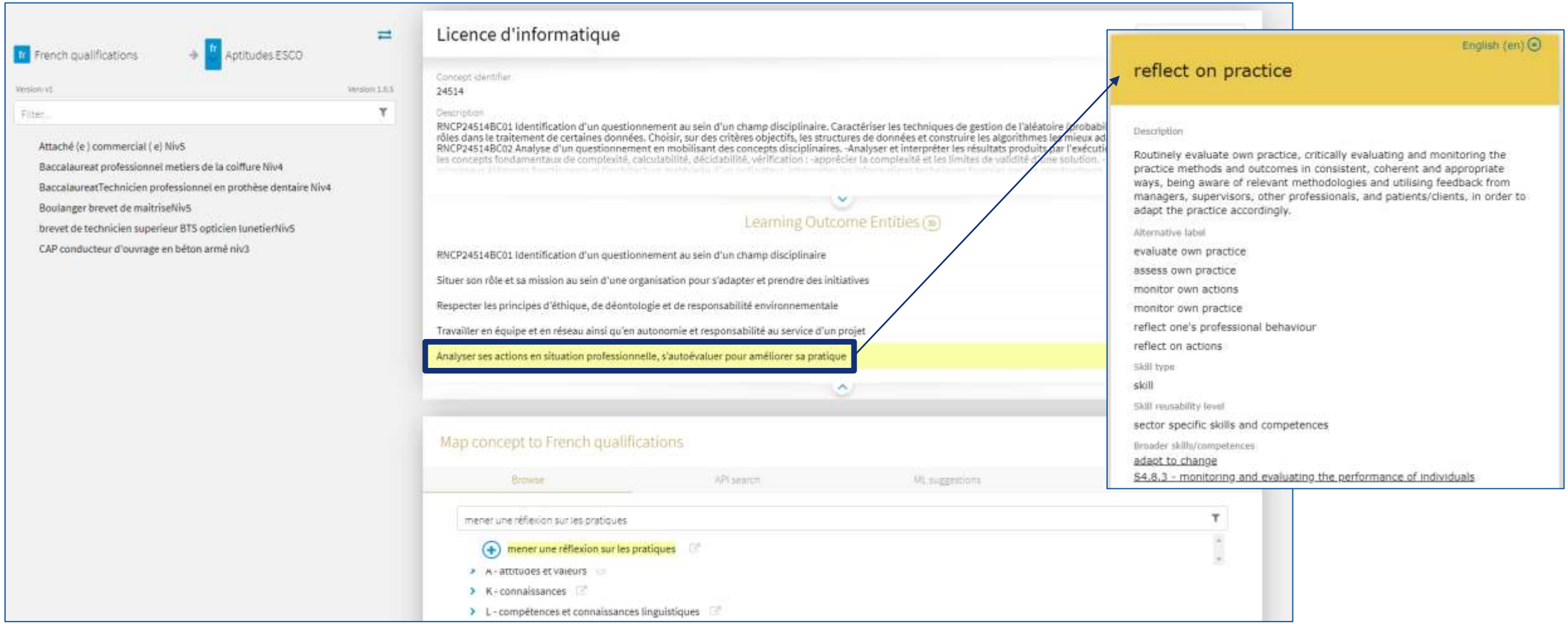
5. Develop algorithms to visualise ESCO

Learning Outcome Linking to ESCO



European Commission

Learning Outcome Linking to ESCO



European Commission

Learning Outcome Linking to ESCO

Machine Learning pipeline in a nutshell

SELECT NATURAL LANGUAGE PROCESSING MODEL

Many different models exist:

- Word2Vec (Google)
- FASTTEXT (Facebook)
- ELMo (AI2)
- Multilingual models

COLLECT DATA

Relevant data are crucial for high quality results. Especially representative data to fine-tune and evaluate the language model.

PRETRAIN NLP MODEL

NLP models come mostly pre-trained.

CONFIGURE NLP MODEL

Each model has many different parameters that can be set to fit the use case

FINETUNE NLP MODEL

The model needs to be trained to produce embeddings that are fit for purpose for the use case and specific text type (vacancy / CV / Qualification)

EVALUATE THE NLP MODEL

The results achieved by the model need to be evaluated against 'ground truth' data.

Implement model in production application

Learning Outcome Linking to ESCO: Early Observations

Learning Outcome Linking to ESCO skills and knowledge: AI model's top 3

what is a computer	implement basic personnel policies and practices	busing, cleaning, and resetting tables	cleaning restroom	ability to prioritise
use a computer	personnel management	clean surfaces	restock toilet facilities' supplies	adjust priorities
computer equipment	advise on personnel management	check dining room cleanliness	maintain the cleanliness of the toilet facilities	exert a goal-oriented leadership role towards colleagues
use ICT systems	manage personnel agenda	manage logs segregation and stacking	conduct cleaning tasks	maintain relationships with stakeholders

National Taxonomies: Early Observations

Germany: mappings to ESCO

German Skill	EN translation	Mapped ESCO skill (EN)	Match type
Freileitungsbau	Overhead line construction	inspect overhead power lines	Close
Freileitungsbau	Overhead line construction	transmission towers	Close
Freileitungsbau	Overhead line construction	repair overhead power lines	Broad
Freileitungsbau	Overhead line construction	install power lines	Narrow

AI model's top 6
repair overhead power lines
repair underground power cables
install low voltage wiring
transmission towers
install power lines
inspect overhead power lines

The AI model mimics the mapping exercise: 4 mapped skills are in the model's top 6

Learning Outcome Linking to ESCO

Machine Learning pipeline in a nutshell

SELECT NATURAL LANGUAGE PROCESSING MODEL

Many different models exist:
- Word2Vec (Google)
- FASTTEXT (Facebook)
- ELMo (AI2)
- Multilingual models

COLLECT DATA

Relevant data are crucial for high quality results. Especially representative data to finetune and evaluate the language model.

PRETRAIN NLP MODEL

NLP models come mostly pretrained.

CONFIGURE NLP MODEL

Each model has many different parameters that can be set to fit the use case

FINETUNE NLP MODEL

The model needs to be trained to produce embeddings that are fit for purpose for the use case and specific text type (vacancy / CV / Qualification)

EVALUATE THE NLP MODEL

The results achieved by the model need to be evaluated against 'ground truth' data.

Implement model in production application

Data for Developing AI Models and Sources

AI needs data to recognise patterns and relations between concepts – more is better!

Learning Outcomes linked to ESCO:

- "*ability to prioritise*" => "*adjust priorities*"

ESCO:

- Occupations and skills: Preferred Terms vs Non-Preferred Terms vs full descriptions
- Skills hierarchy: "*evaluate practice in psychotherapy*" is more narrow than "*reflect on practice*"
- Relations: skill-skill, occupation-skill, occupation-occupation

Mappings between national taxonomies and ESCO

Other taxonomies: O*NET, UKSOC, SSOC, ISCO

Qualifications in the QDR:

- ISCED-F for learning opportunity
- Learning outcomes vs qualification titles

Data for Developing AI Models and Sources

Wanted: data that represent/contain **relations** between concepts or between free text

Relation Type	Phrase 1	Phrase 2	Ex Data Source
Synonyms	operative for groundworks	groundworker	ESCO
Alternatives	talent aquisition manager	recruitment manager	ESCO
Description	babysitter	they provide short-term care services to children on the premises of the employer	Taxonomies, Online Job Vac
Mapping	looking for an experienced breakfast chef!	breakfast cook	Online Job Vac, Job Search Engines
Multilang. mapping	Freileitungsbau	install overhead power lines	Mappings to ESCO
Broader/Narrow	humanties teacher	history school teacher	Taxonomies
Related	statistical programmer	data scientist	CVs
Qual – Occup/LO	master in artificial intelligence	machine learning engineer	QDR, CVs
Skill – Job Title/Occupation, Skill – Skill	manage a team of sales representatives	sales manager	Job search user queries, Click logs, ESCO
Seniority	junior software developer	software application development manager	CVs
Related duties	monitor IT security systems	audit different aspects of the security program to ensure compliance	Online Job Vac, ESCO, O*NET
Tagged	renewable energy	green skill	
Coded	greenhouse types	ISCED-F – 0812	ISCED-F
Coded	senior radiocommunications technician	ANZSCO – 313211	ANZSCO

Data for Developing AI Models and Sources

Multilingual relations:

- AI methodology is usually first developed for English, then extended to other languages
- Minimise language specific components in the AI pipeline

Different length:

- We are not only looking for relations between short phrases.
- Long pieces of text can be processed as well

Structured vs Unstructured data:

- E.g. highly structured online job vacancies (skills/duties/qualification/…) vs flat online job vacancies (OJV)

Real-world data: Taxonomies like ESCO are clean, we also aim for noisy data like OJV

Geographical coverage: it is important to have data from all member states

Learning Outcome Linking to ESCO

Machine Learning pipeline in a nutshell

SELECT NATURAL LANGUAGE PROCESSING MODEL

Many different models exist:

- Word2Vec (Google)
- FASTTEXT (Facebook)
- ELMo (AI2)
- Multilingual models

COLLECT DATA

Relevant data are crucial for high quality results. Especially representative data to finetune and evaluate the language model.

PRETRAIN NLP MODEL

NLP models come mostly pretrained.

CONFIGURE NLP MODEL

Each model has many different parameters that can be set to fit the use case

FINETUNE NLP MODEL

The model needs to be trained to produce embeddings that are fit for purpose for the use case and specific text type (vacancy / CV / Qualification)

EVALUATE THE NLP MODEL

The results achieved by the model need to be evaluated against 'ground truth' data.

Implement model in production application

Evaluating Artificial Intelligence Models

German Skill	EN translation	Mapped ESCO skill (EN)	Match type
Tierphysiologie	Animal physiology	physiology of animals	Exact
Tierphysiologie	Animal physiology	neurophysiology of animals	Broad
Tierphysiologie	Animal physiology	analyse animal locomotion	Close
Tierphysiologie	Animal physiology	use osteopathic techniques to improve health of animals	Close

AI model's top 9
physiology of animals
examine animals
perform veterinary diagnosis
anatomy of animals
assess animal's condition
conduct experiments on animals
animal biology
fundamental veterinary sciences
neurophysiology of animals

How good is the validated relation: *"use osteopathic techniques to improve health of animals"?*

Evaluating Artificial Intelligence Models

Evaluating the quality of a model is a non-trivial task:

- Evaluating a model should be based on **objective criteria**, e.g.:
  - Number of correctly mapped skills to ESCO
  - Percentage of correctly predicted ISCED-F codes
  - Percentage of correctly matched job title synonyms

- Creating a '**ground truth data set**' can be time consuming but is important:
  - Diverse set of examples for which we know the correct result
  - Not all model errors are equally bad

- Ground truth data and objective criteria enable us to do **benchmarking** of results

#ESCO_EU

empl-esco-secretariat@ec.europa.eu


THANK YOU