



---

**AI-based tools for  
patent classification**

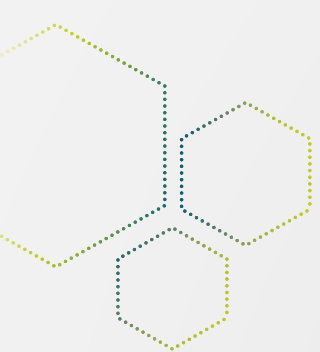
---

WIPO SCP/35 – Sharing session on the use of AI for patents – 17<sup>th</sup> October 2023

**Samir Ghamri-Doudane – Head of Lab INPI (France)**

**inpi**





# SUMMARY



## AI FOR...

- Patent Pre-classification
- Patent Classification
- Other Applications

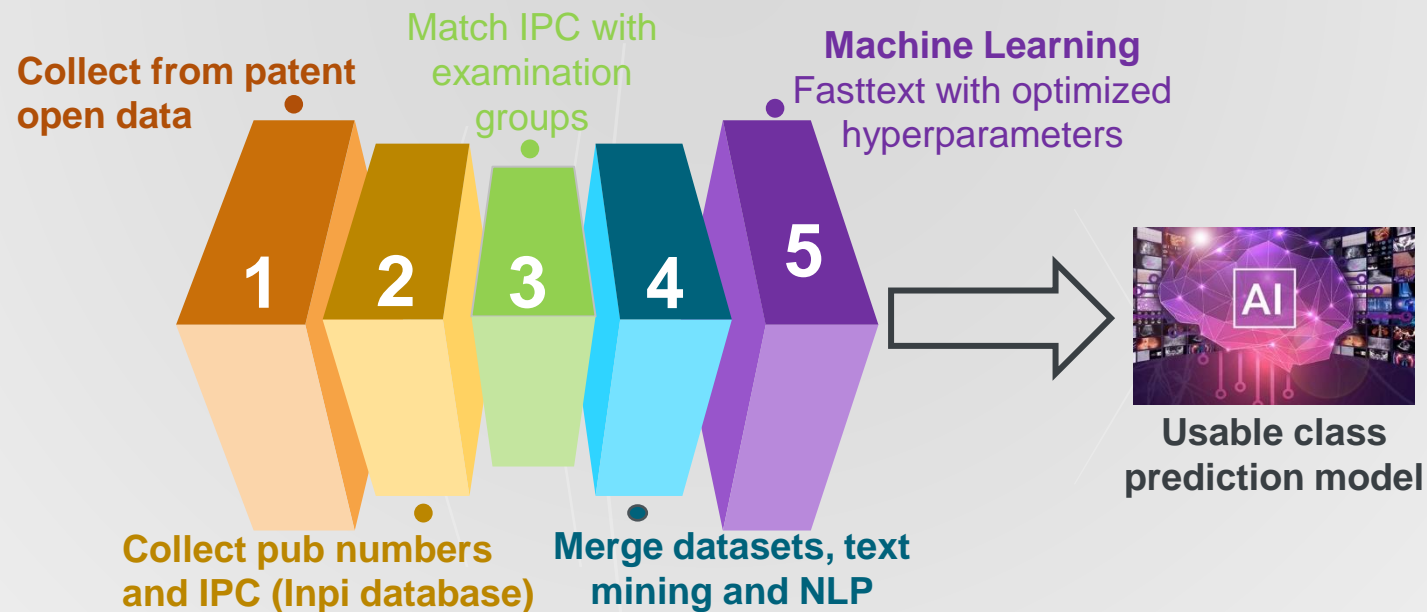




# PATENT PRECLASSIFICATION

## First model development using AI

- ▶ **Focus on patent distribution** among examination teams with automation
- ▶ **Data collection** with 150 000 patents published by INPI over a 10 years period
- ▶ **Natural language processing** to prepare and pre-process data
- ▶ **Supervised learning** with the adaptation of a state-of-the-art model

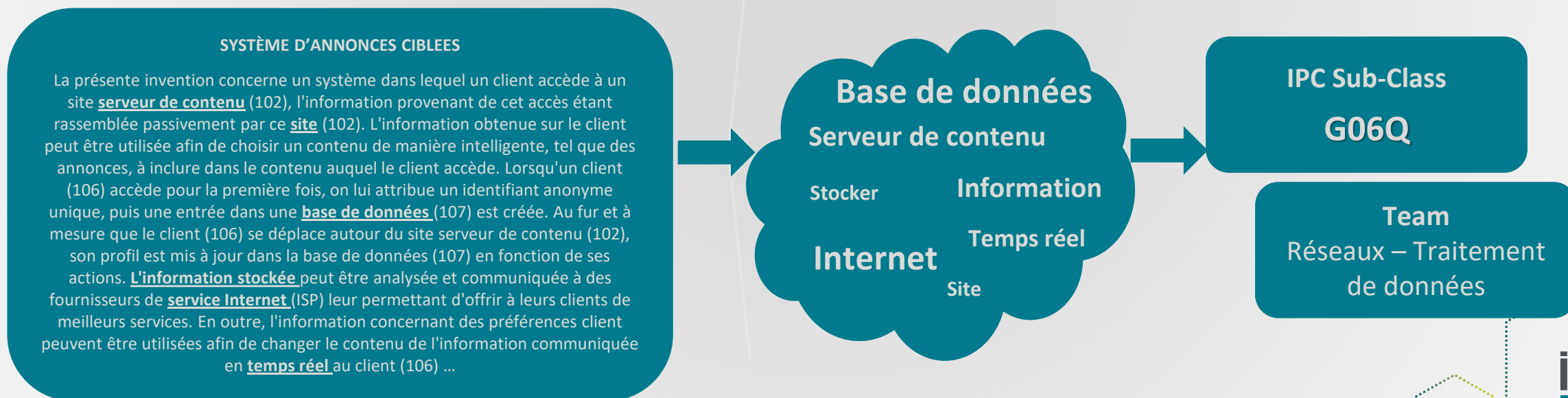


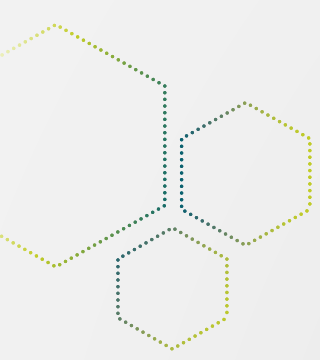
WIPO FOR OFFICIAL USE ONLY

# PATENT PRECLASSIFICATION

## Successful automation

- ▶ **In production** since 2019 with regular use by the patent department replacing manual dispatching
- ▶ **Recent model improvement** use of a Bi-LSTM to predict technology sub-class before assignment
- ▶ **Good accuracy**, around 80% when first deployed and reached 92% with the new model
- ▶ **Significant time savings** for service managers estimated at 20 to 30 hours per week in total





# PATENT CLASSIFICATION

## A complex task ...

### IPC: A 5-level classification hierarchy

<b>G</b>	Physics
<b>G06</b>	Computing; Calculating; Counting
<b>G06F</b>	Electric digital data processing
<b>G06F 8/00</b>	Arrangements for software engineering
<b>G06F 8/20</b>	Software design

- ▶ An **I**nternational **P**atent **C**lassification to characterize the addressed technological domain
- ▶ A multi-label classification with one primary code and several secondary codes assigned for each patent
- ▶ A fine-grained classification with more than 60 000 subdivisions at lower level!





# PATENT CLASSIFICATION ... requiring sophisticated models

- ▶ **Focus** on extreme multi-label text classification models to move toward automation
- ▶ **Feasibility study** by evaluating state-of-the-art approaches on different levels of the IPC
- ▶ **Collaboration** with a research institute on this cutting-edge technology

## XML-CNN

A family of convolutional neural networks that is tailored to extreme multi-label text classification.

<http://nyc.lti.cs.cmu.edu/yiming/Publications/jliu-sigir17.pdf>

## Parabel

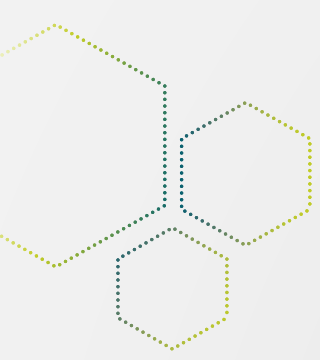
Partitioned Label Trees for extreme multi-label learning and classification.

<http://manikvarma.org/code/Parabel/download.html>

## BERT for Patent

Leveraging the BERT model for patent multi-label Classification.

<https://cloud.google.com/blog/products/ai-machine-learning/how-ai-improves-patent-analysis>



# PATENT CLASSIFICATION

## Data is key

- ▶ **Significant work** achieved to have the right quality and quantity of patent data
- ▶ **Data augmentation** using different sources is an important subsequent step

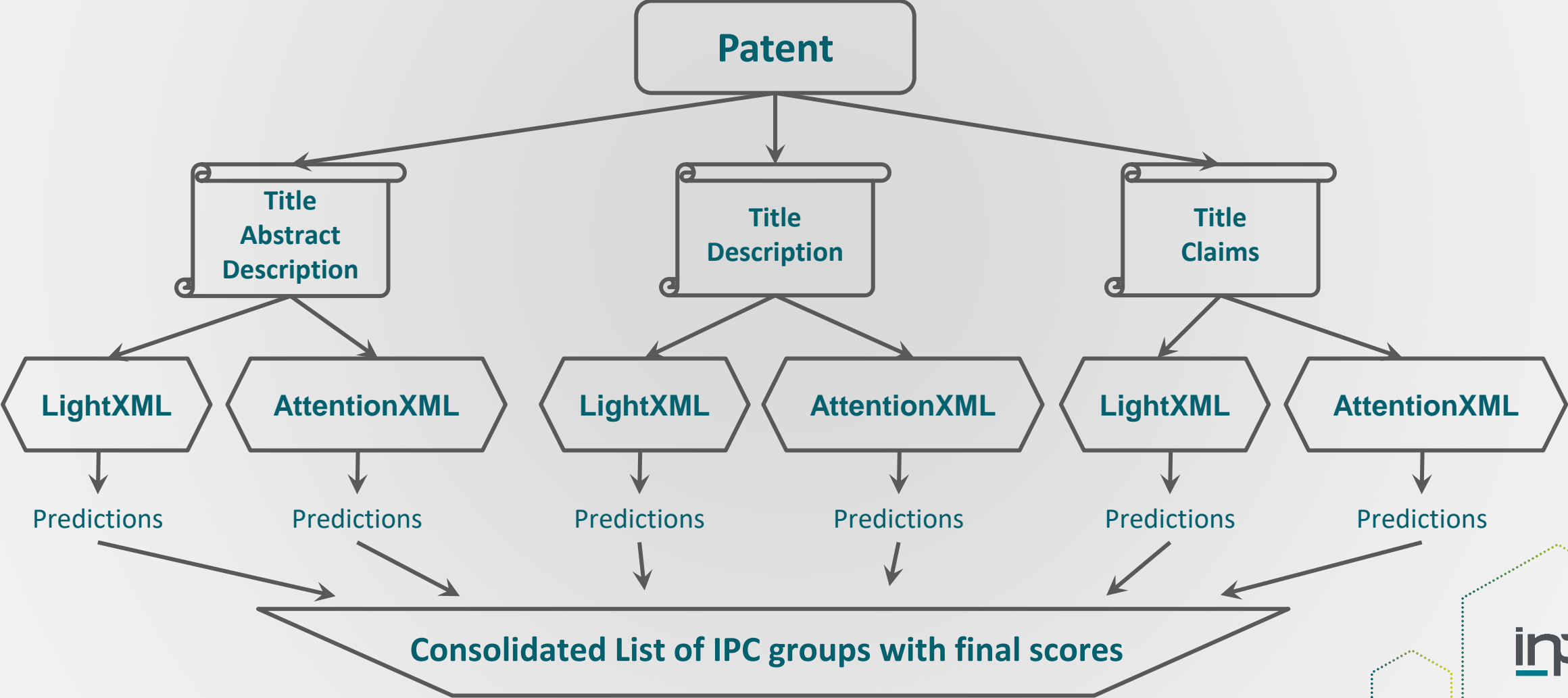
### Dataset source and language

Patent Section		EPO (en)	EPO (fr)	INPI (fr)
	Title	2 155 349	2 155 346	296 270
	Abstract	1 029 261	605 770	295 421
	Description	1 849 071	70 814	296 216
	Claims	2 155 361	1 125 999	291539

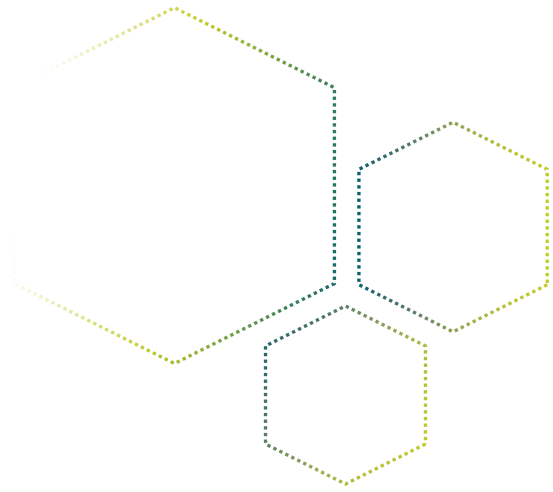


# PATENT CLASSIFICATION

With a combination of models







# Thank you



[www.inpi.fr](http://www.inpi.fr)



+33 (0)1 56 65 89 98



INPI France

[hmouzoun@inpi.fr](mailto:hmouzoun@inpi.fr) – [sghamridoudane@inpi.fr](mailto:sghamridoudane@inpi.fr)

**inpi**

