

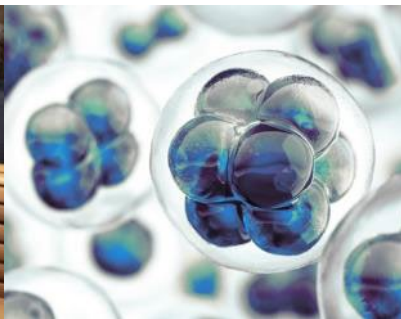


Europäisches
Patentamt
European
Patent Office
Office européen
des brevets

AI applied to Classification Use Cases at the European Patent Office



Alexander Klenner-Bajaja



Head of Data Science Department



IPC Union - Committee of Experts: Fifty-Fourth Session

Strategic Plan 2023

- <https://www.epo.org/about-us/office/strategy.html>

Our goals for 2023
The Strategic Plan comprises five goals, each of which has several key initiatives.

Goal 1
Engaged and empowered
1
Read more

Goal 2
Digital transformation
2
Read more

Goal 3
Master quality
3
Read more

Goal 4
Partner for positive impact
4
Read more

Goal 5
Secure sustainability
5
Read more

Programme
Artificial Intelligence & Blockchain

Programme breakdown

Artificial Intelligence Essentials

Computer Vision

- Patent figures
- Facsimiles parsing e.g. tables
- Understanding figure content

Natural Language Processing

- Managing patent language
- Patent syntax in general
- e.g. claim syntax specifically

Machine Translation

- Specific to our business needs
- Build on patent content
- Proven to improve PreSearch/PreCla

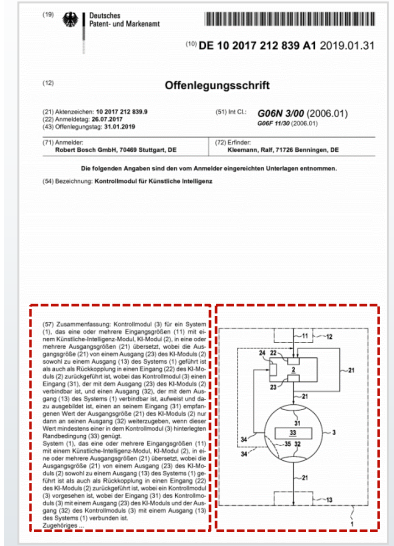
Blockchain

KPI3.2 Quality of search and examination (%) (Internal Audit)

KPI3.2 Quality of search and examination (%) (Internal Audit)

KPI3.2 Quality of search and examination (%) (Internal Audit)

KPI3.2 Quality of search and examination (%) (Internal Audit)



EPO is focusing on



AI based Pre-Classification

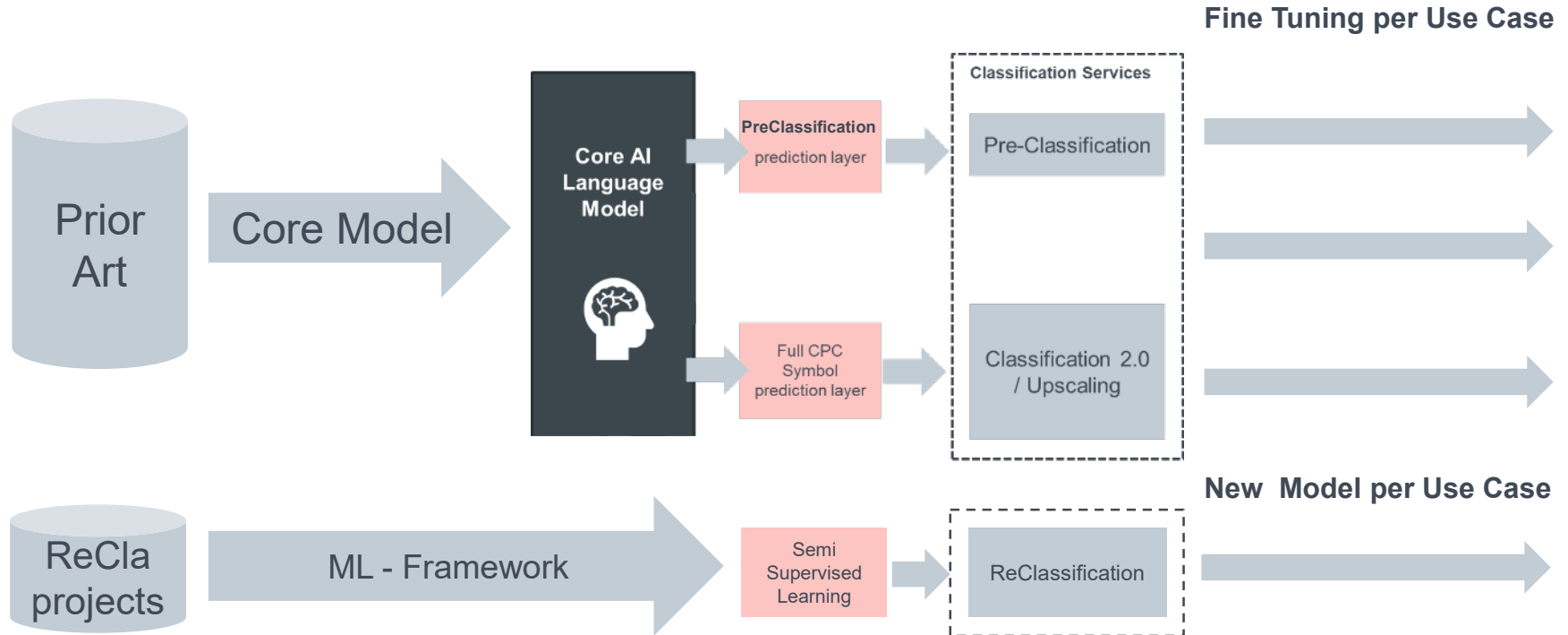


AI based Full-Classification



AI supported Re-Classification

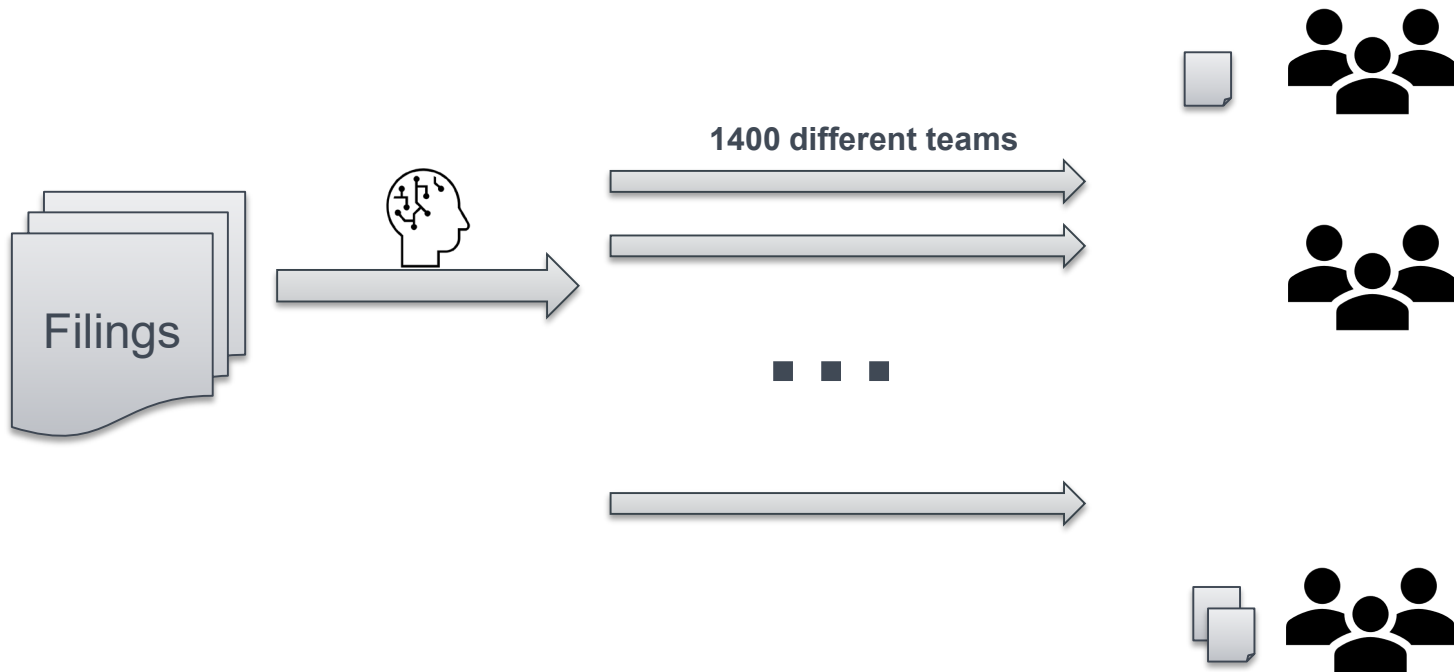
Natural Language Processing Project



Pre-Classification

Patent Application to the correct team

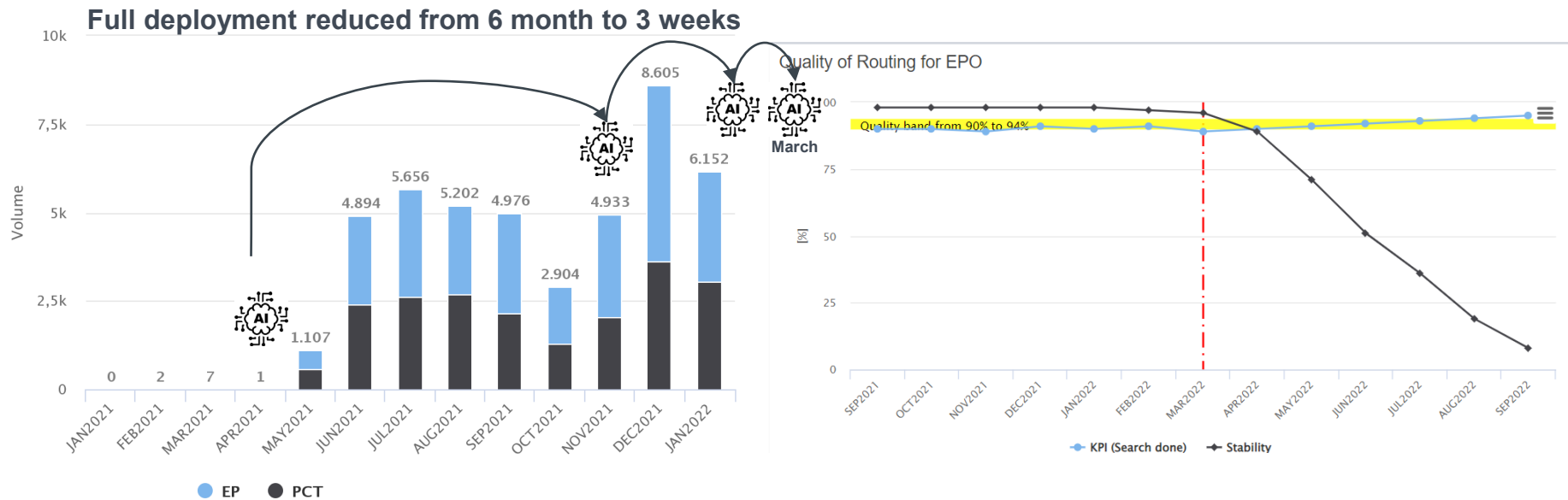
Pre-Classification



1 400 possible ranges (labels)

800.000 training documents from 2013-2019 (EPO pre-classified, sliding window)

Pre-Classification – Business KPI – Prospective Evaluation



1 400 possible ranges (labels)

~600k training documents from 2013-2023 (EPO pre-classified, sliding window)

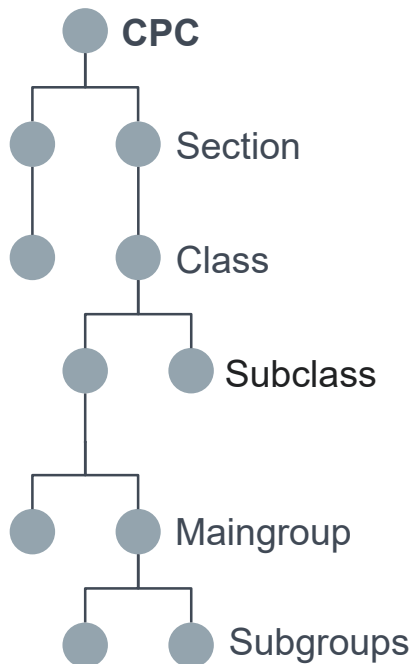
(Feedback received)

Human Quality Control built in

Full CPC Classification

Finding the correct CPC symbols

Classification – CPC classification on simple family level

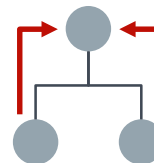


Data available

~250 000 symbols in CPC

Aggregate up until

- a) Main group or
- b) Until 50 docs are collected



117k symbols after aggregation
→ **33% of CPC**

Training on 7M
Test on 1M

800k families from 2021:

Unique symbols coverage:
(60.4%)

Total symbols coverage:
(95.7%)

Version without aggregation
Covering 100% in training.

Classification – Retrospective Performance

- Soft Accuracy@5 = 0.81

- Recall@10 = 0.51



In 80% we find at least 1
examiner given symbol in
the top 5

On average we find 50%
of all given symbols in the
top 10

- Extremely good given the complexity of the task (3-5 out of 117k)
- The data is challenging as well → see next slide

Tool available to all examiners: Canopee

The screenshot shows the Canopee web interface in Google Chrome. The browser address bar shows the URL: [https://batavia.internal.epo.org/ansepa/master/canopee/#\(\(browser.details,row,60\),\(cromo.class-alloc,row,30\),col,59\)/info](https://batavia.internal.epo.org/ansepa/master/canopee/#((browser.details,row,60),(cromo.class-alloc,row,30),col,59)/info). The interface is divided into several sections:

- Class Browser:** Shows a list of classes for the scheme "CN1067300C - Confirmed". The selected class is **A46D1/00** with the description: "Bristles; Selection of materials for bristles (making artificial bristles D01D, D01F)". Other visible classes include B44D3/00, Y10T29/00, and Y10T29/49.
- Detail window:** A large empty area on the right with the text "Click on the icons more".
- Cromo:** A table showing publication information for the class allocation.
- Class Alloc - CN1067300C:** A table showing the distribution of symbols across different categories.

Publication	Due date	Pub. date	Vis
CN1185760A	-	24/06/1998	

AI suggestion	Confirmed	IPC
B05C17/00 (69.6%)	A46D1/00 (EPO)	A46D1/00
A A46B2200/202 (51.8%)	B44D3/00 (EPO)	B05C17/00
B44D3/00 (39.3%)	A Y10T29/49925 (EPO)	B43K1/00
B05C17/02 (22.6%)	A Y10T29/49938 (EPO)	B44D3/00
	A Y10T29/49964 (EPO)	

AI suggestion

- B05C17/00 (69.6%)
- A A46B2200/202 (51.8%)
- B44D3/00 (39.3%)
- B05C17/02 (22.6%)
- B05C17/0207 (10.1%)
- B05C17/10 (9.1%)
- B25G1/102 (8.8%)
- B05C17/002 (7.6%)
- B05C17/0205 (7.5%)
- A A46B2200/205 (5.7%)

Re-Classification

Reacting to CPC revisions

Reclassification: Data Science perspective

- Provide a **framework** in which each reclassification work can be **Semi-supervised** classification task
- Hierarchical **multi-label** text-classification
- Classification with **few training examples** (~10 documents per target symbol)
- Reclassification **workload** at EPO:
 - 100k-200k documents/year
 - 50-100 reclassification projects, high variability
- Goal: High quality (90%+) while minimizing manual effort

The reclassification Tool

Projects > RP0676_SA32

RP0676_SA32

EPUS NPL IRWL - Created on 23/12/2021 15:42:38

[Download report](#) [Finish Reclassification](#)

[Project progress](#)

Overview at Project Level - EPUS IRWL [Help on AI assistant](#) [Intellectual recla at project level](#)

Required Quality Level:

Total number of documents: 1059

Documents fully reclassified (for all source groups): 18






Docs. reclassified by exa. for at least one source group: 18

Docs. reclassified by AI for at least one source group: 0

Supported by AI: 824

Not supported by AI: 235

Overview per source group of RCL

Source group	Type	Transfer type	Nbr. of RCL destination groups	AI stage	Total number of documents	Docs. not in status "recla. completed"	Reclassified by examiner	Reclassified by AI	Supported by AI	
B29C70/00		C	INT	3	Not started	10	10	0	0	6
B29C70/30		C	INT	3	Not started	10	10	0	0	9
B29C70/46		C	INT	2	Not started	10	10	0	0	7
B29C70/54		C	INT	3	 QC 1	1022	1004	18	0	796
B29D99/0003					Not started	10	10	0	0	8

Intellectual recla for this group
AI recla for this group

All 5 loaded.

What is cooking?

- Auto-Classification
 - EP-BERT to EP-RoBERTa
 - 100% CPC coverage
- Pre-Classification
 - Increasing of volume (coverage of all filing types)
 - EP-BERT to EP-RoBERTa
- Auto-Classification as Direct File Allocation System (rollout in 23)

EPO is focusing on



AI based Pre-Classification



AI based Full-Classification



AI supported Re-Classification



Thank you.