

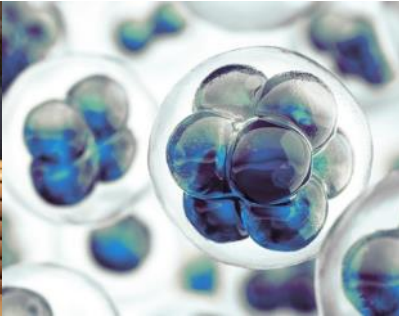


Europäisches  
Patentamt  
European  
Patent Office  
Office européen  
des brevets

# AI applied to classification use cases



Alex Klenner-Bajaja



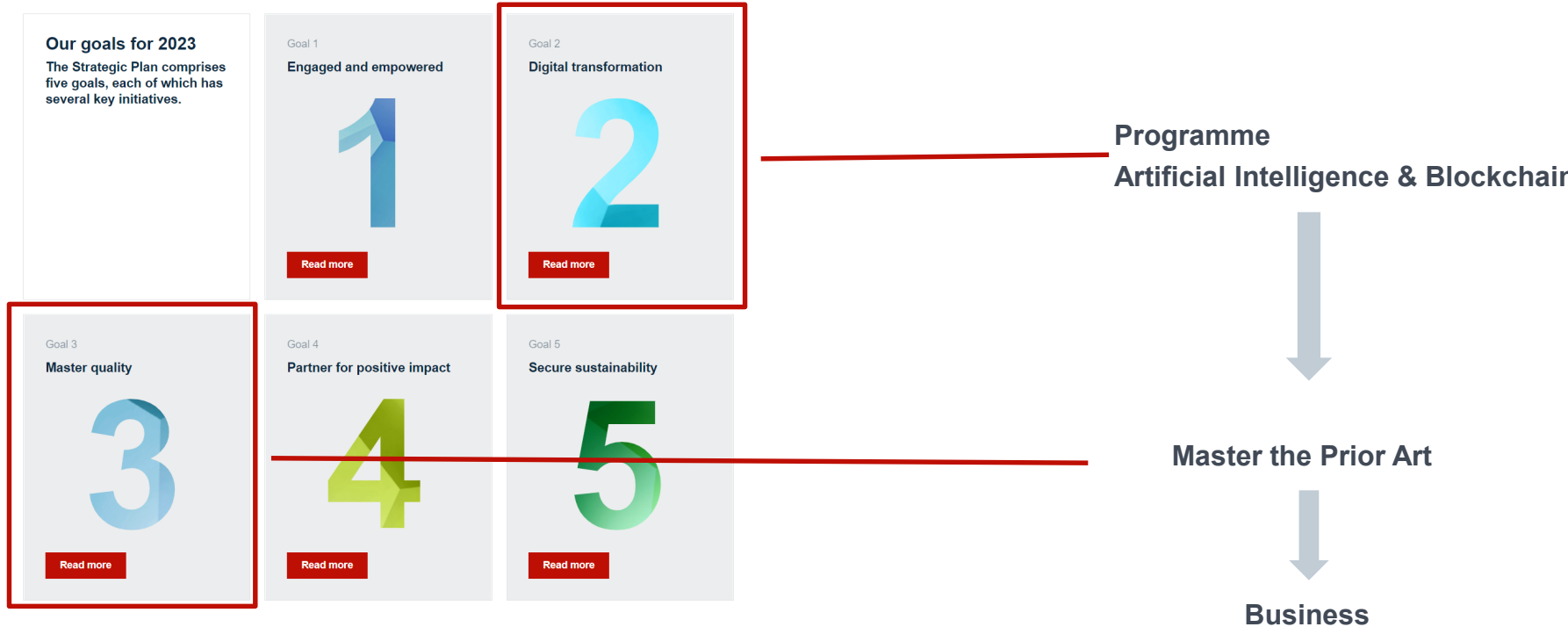
Head of Data Science Department



25-02-2022

# Strategic Plan 2023

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



# 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

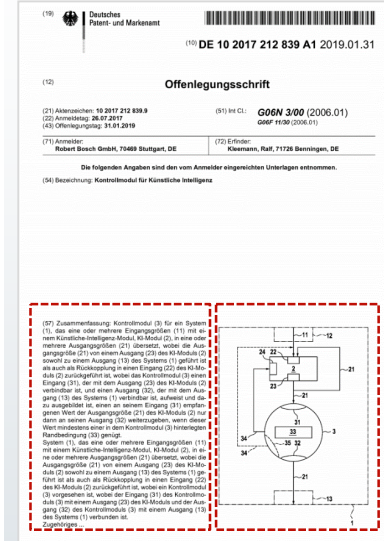
### Blockchain

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**

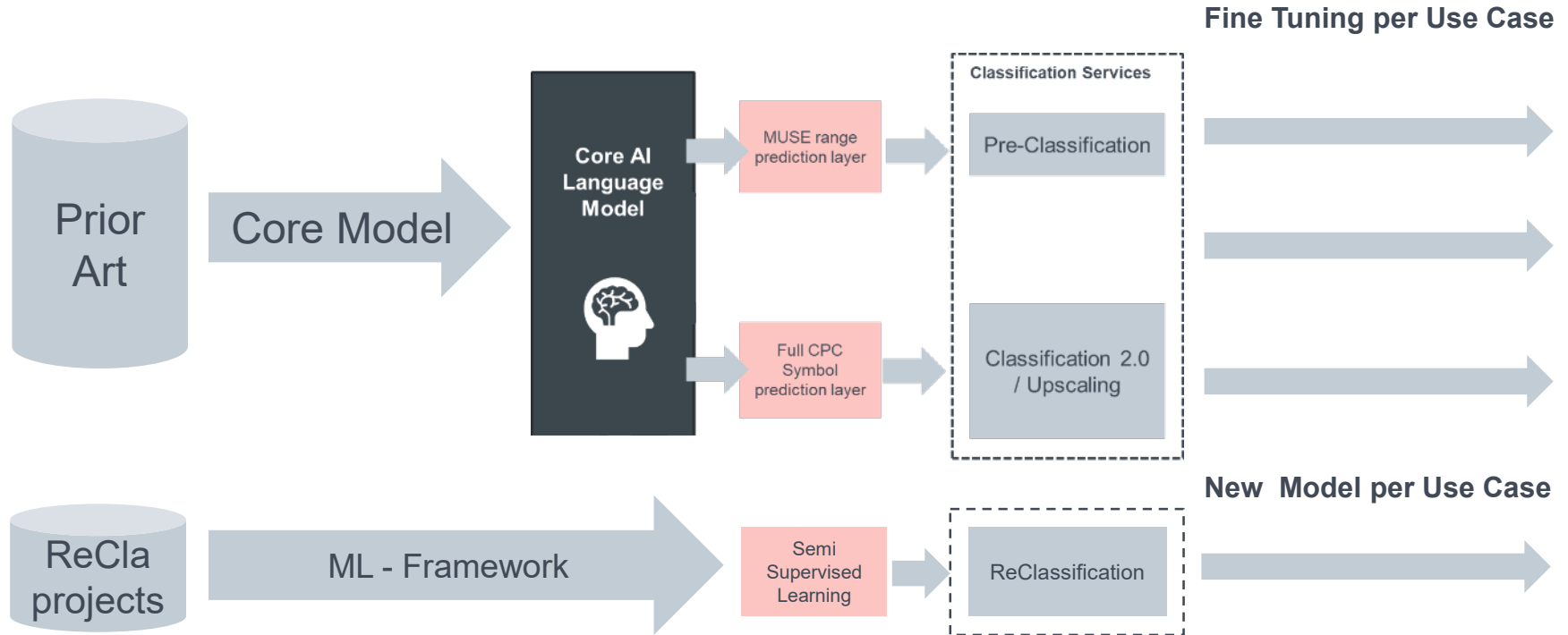


**AI based Full-Classification**



**AI supported Re-Classification**

# Natural Language Processing Project



# Some insights into EP-Bert inner workings

Both	EP-BERT	BERT
<i>the, of, and, in, to, was, he, is, as, for, on, with, that, it, his, by, at, from, her, she, you, had, an, were, but, be, this, are, not, my, they, one, which, or, have, him, me, first, all, also, their, has, up, who, out, been, when, after, there</i>	<i>burners, transfecting, isobuty, intr, ected, tumours, directionality, implanting, disassoci, microw, allele, sectioned, hortici, tetrazole, interferogram, unb, varn, imprinting, proportioning, hydrocarbylene, carbox, decoupled, bts, parenteral, perm, inb, parad, nsc, centi, perhydro, atgg, inamide, homopolymer, decrypting, suspensions, encapsulated, intranasal, unlatched, glycosidic, hexafluoride, phenolic, pyramidal, metabis, homod, dither, flattening, exfoliation, hape, oterol, lytic</i>	<i>ushered, blaming, nais, statehood, cadillac, viola, ska, patsy, sorority, rothschild, soundtrack, 1920s, distracting, tnt, touring, blessed, counselor, bial, tation, inevitably, nintendo, cornelius, soviets, oge, stoke, parry, evolve, evil, africa, marilyn, farrell, subway, haggard, costing, laise, styx, booths, paces, scientists, crater, gale, spielberg, ashi, replied, exclaimed, patrons, thirteen, carnegie, zombies</i>

(almost 1/3 = 10k from 32k)

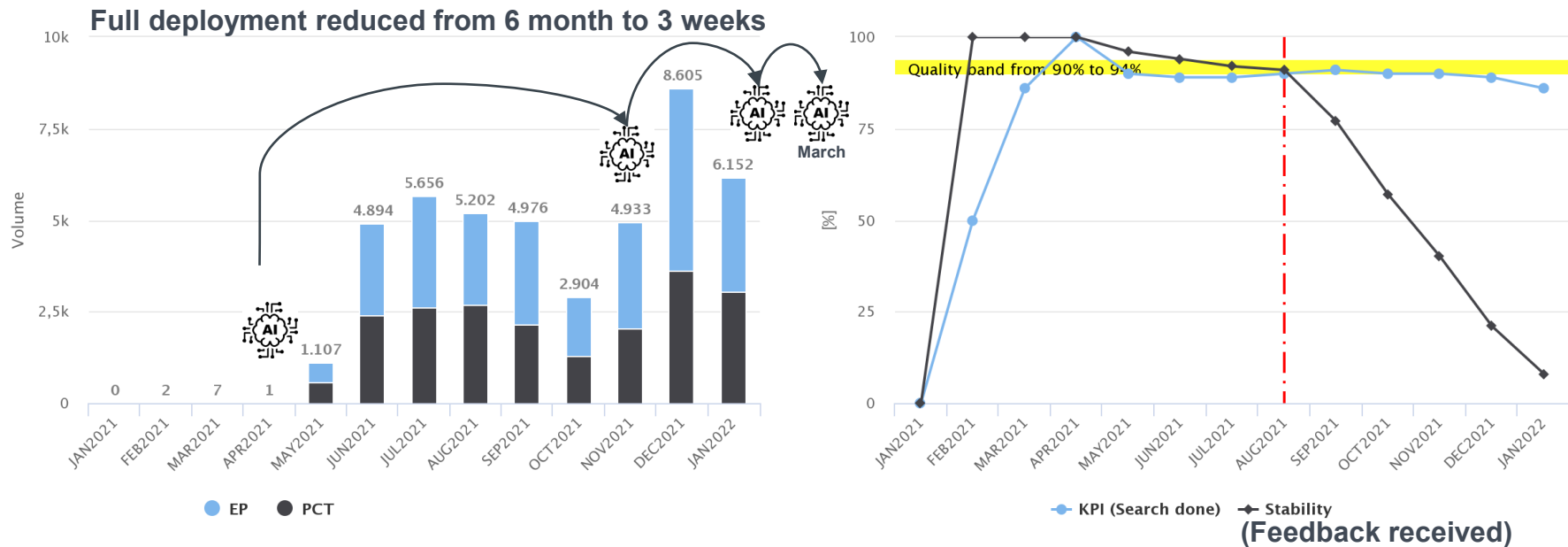


**Huge impact on sequence length for prediction**

# **Preclassification**

**Patent application to the correct team**

# Preclassification – Business KPI – Prospective Evaluation



**1 400 possible ranges (labels)**

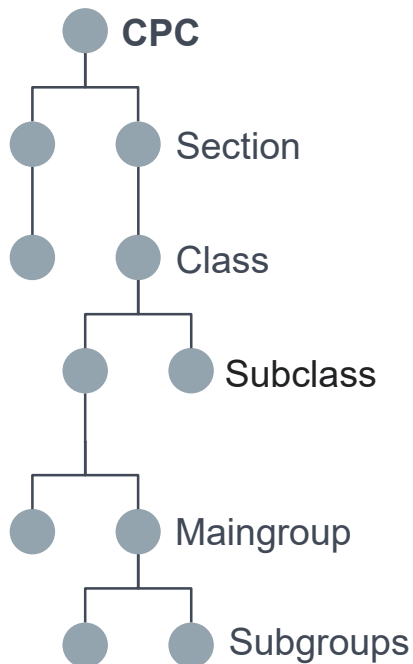
**~600k training documents** from 2013-2019 (EPO pre-classified, sliding window, latest until 14-21)



# **Full CPC Classification**

## **Finding the correct CPC symbols**

# Classification – CPC classification on simple patent family level

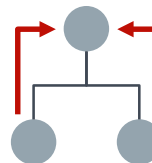


Data available

~270 000 symbols in CPC

Aggregate up until

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



**81k symbols after aggregation**

Training on 3.3M

Test on 0.83M

# 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

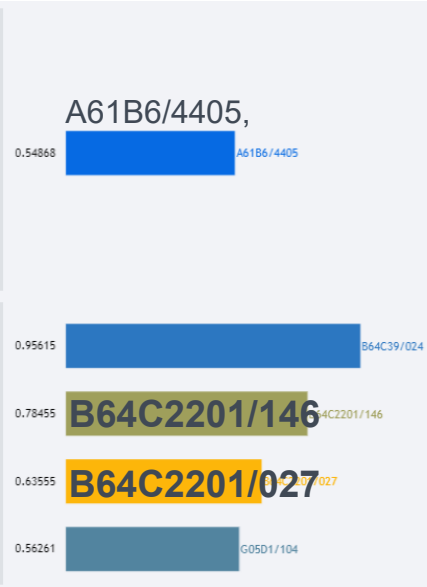
On average we find 50%  
of all given symbols in the  
top 10

- Extremely good given the complexity of the task (3-5 out of 80k)
- The data is challenging as well → see next slide

# Classification – An example

13 The X-ray imaging system may comprise at least one data storage configured to store and/or storing computer-executable instructions, which when executed by the controller, one or more of the sub-controllers, and/or one or more of the processors instruct the imaging system to carry out any of the functions described above and in the following with respect to the X-ray imaging system, with respect to a component and/or a part of the X-ray imaging system, and/or with respect to any other aspect of the invention, such as the X-ray imaging arrangement and the method. The data storage may comprise a plurality of data storage devices, e.g. arranged in different parts and/or components of the imaging system. Accordingly, at least one data storage device may be arranged in the unmanned aircraft, a further unmanned aircraft, the X-ray detector and/or the X-source.

14 The at least one unmanned aircraft may refer to an unmanned aerial vehicle, a drone, a quadcopter, and/or any other type of mobile module configured to fly. Accordingly, the at least one aircraft may comprise at least one propelling unit, a drive and/or a motor enabling the at least one unmanned aircraft to fly and/or move three-dimensionally. The at least one unmanned aircraft may be at least in part remotely controlled. The at least one unmanned aircraft may comprise a communication interface, a communication module and/or a communication element to communicate, particularly bidirectionally communicate, with another device, such as e.g. the control station and/or the control device.



CPC confirmed invention information

**A61B6/4405**, A61B6/4452,  
A61B6/548, A61B6/587,  
A61B6/588,

G05D1/0094

CPC confirmed additional information

G01N23/04 (x-ray)

**[...] unmanned aircraft may refer to an unmanned aerial vehicle [...] remotely controlled**

**B64C2201: Unmanned aerial vehicles; Equipment therefor**

**/027 Flying platforms**

**/146 Remote controls**

# A glimpse of the tool (prototype)

The screenshot shows the Canopée tool interface in a Google Chrome browser. The main window is titled "Class Browser" and displays a search bar and a list of classification symbols (A through G) with their corresponding titles and document counts. The "Class Info" panel on the right shows details for class C05, including its definition and notes. The "Class Alloc" table at the bottom shows a list of documents with columns for Unreviewed, Confirmed, CPC other off..., AI Suggestions, IPC, and Circulation. The "AI Suggestions" column is highlighted with a red box, and a red arrow points to the "AI Suggestions" header. Another red arrow points to the "AI Suggestions" column in the table, and a third red arrow points to the "AI Suggestions" column in the "AI Suggestions" panel.

Doc name	Due date	Pub. date	Visit	Unreviewed	Confirmed	CPC other off...	AI Suggestions	IPC	Circulation
CN105235544A	13/03/2016	12/01/2016	Clipo		B01J23/18	B01J23/18	B01J23/18	B01J23/18	
US2019293502A1	26/11/2019	25/09/2019	CPC		B01J23/22	B01J23/22	B01J23/20	B01J23/20	
US2019293504A1	26/11/2019	25/09/2019	CPC		B01J37/08	B01J37/08	B01J23/22	B01J23/22	
					C07C253/24	C07C253/24	B01J23/847	B01J23/847	
					C07C253/26	C07C253/26	B01J37/08	B01J37/08	

The "AI Suggestions" panel displays a list of classification symbols: B01J23/18, B01J, B01J23/22, B01J23/847, and B01J37/08. A red box highlights the "AI" button next to the B01J symbol. A red arrow points from the text "Allocate symbol as 'I' or 'A'" to the "AI" button. Another red arrow points from the text "Delete" to the "Delete" button next to the B01J symbol.

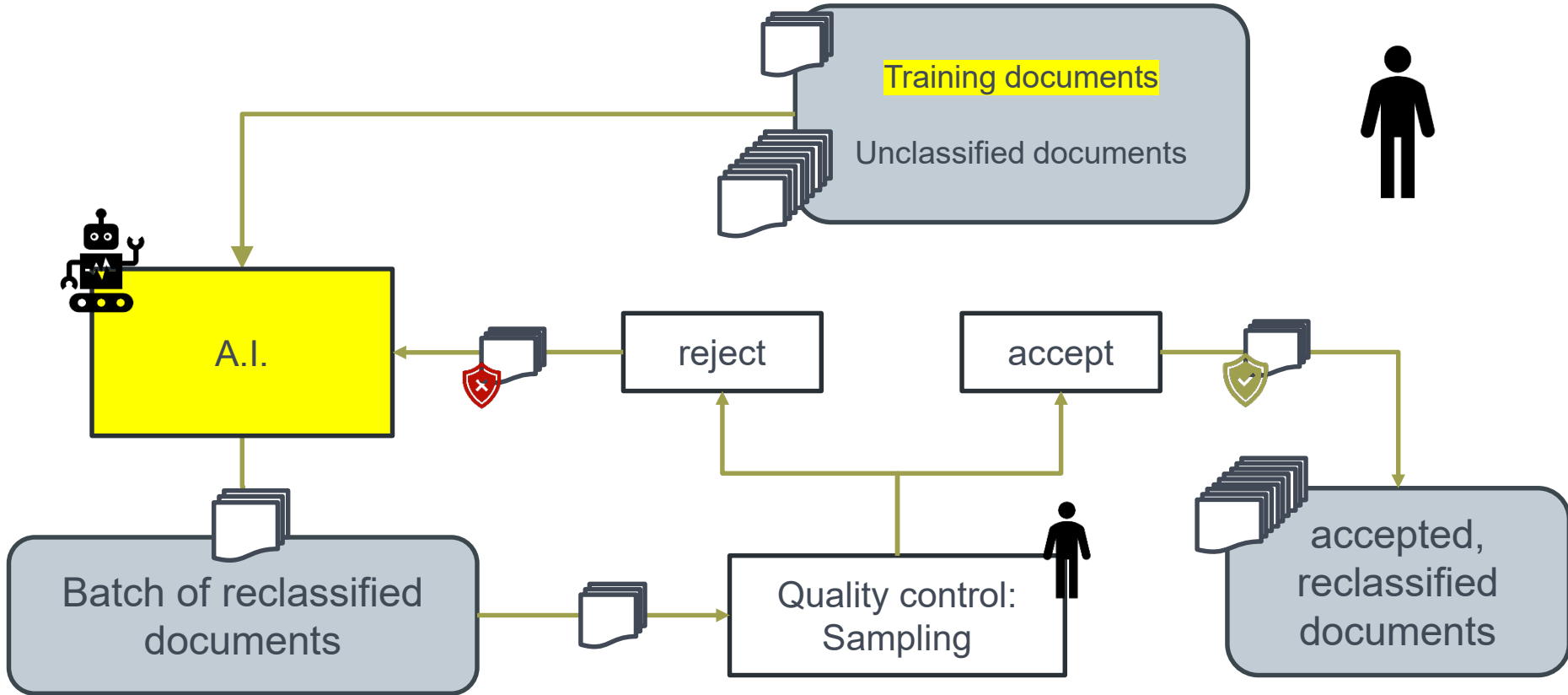
Allocate symbol as "I" or "A"

Delete

# **Reclassification**

## **Reacting to CPC revisions**

# The reclassification task: A.I.



# Reclassification: data science perspective

- Provide a **framework** in which each reclassification work can be **Semi-supervised** classification task (inference data known)
- 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



# A glimpse of the 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		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.

# EPO is focusing on



**AI based Pre-Classification**



**AI based Full-Classification**



**AI supported Re-Classification**



**Thank you**