

From the ML Model to Practice

Case Study on NLP-based Decision-Making on the Eligibility of Security Prospectuses



Maximilian König
AI Solution Architect



Bernd Rusitschka
AI Expert in DG
Markets



Janek Blankenburg
AI Application Engineer



Philipp Rothhaar
Expert in DG Markets

In collaboration with further colleagues from DG Markets and Prof. Christian Hänig and Serhii Hamotskyi from Anhalt University of Applied Sciences

Agenda

Status quo ante Deciding the Eligibility of Securities' Prospectuses

Training a model Proof of Concept with a fine-tuned model

Integration ... of the model into the business process

Learnings ... from the process

Status Quo Ante

Deciding the Eligibility of Securities' Prospectuses

Status Quo Ante

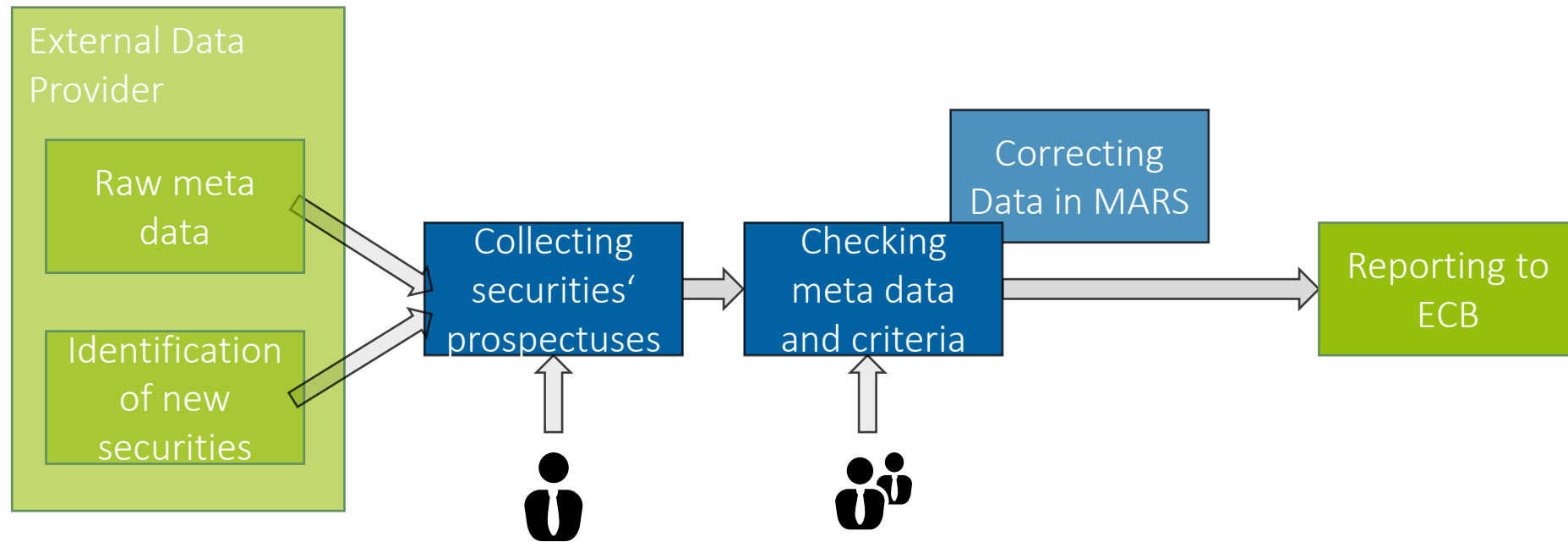
Deciding the Eligibility of Securities Prospectuses



- NCBs report daily new **eligible marketable assets** to ECB, which collects them into **EADB (Eligible Assets Database)**
- Checking a security / asset for eligibility is based on harmonized criteria (Guideline (EU) 2015/510)
- The reporting contains the **eligible assets** as well as related **meta data**
- Several eligibility criteria are established based on a security's prospectus
 - So far this is achieved by manually checking / reading the prospectuses in a four-eyes principle
- At Deutsche Bundesbank (BBk) the (BBk-made) application MARS is used for collecting the securities' data and reporting them to ECB

Status Quo Ante

Process Flow



Manual assessment is time-consuming and repetitive

Training a model

Proof of concept with a fine-tuned model

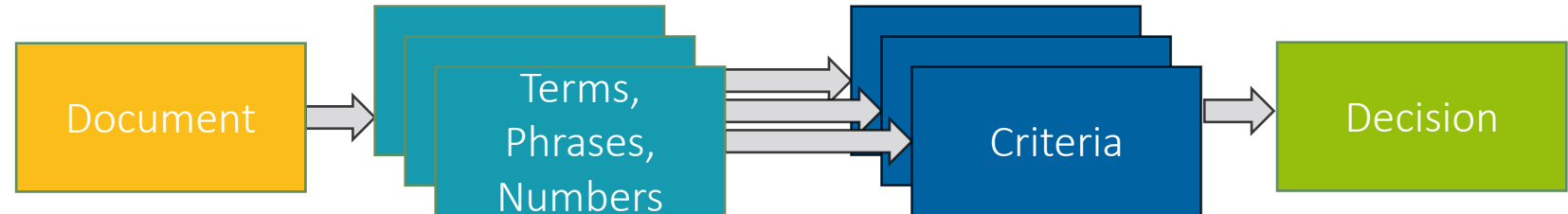
Research Project Automatic Annotation

Proof of Concept using NLP

§ 5
(Status)

Die Schuldverschreibungen begründen ~~x status_nicht_nachrangig_bevorrechtigt~~ nicht besicherte und nicht nachrangige Verbindlichkeiten der Emittentin. Bei Emission handelt es sich bei den Schuldverschreibungen um bevorrechtigte Schuldtitel (Senior Preferred Schuldverschreibungen), die nicht den durch § 46f Absatz 5 in Verbindung mit Absatz 6 KWG gesetzlich bestimmten niedrigeren Rang haben.

- Task at hand: Identifying in PDF-Documents a given number of terms, phrases, numbers etc. that form the basis for the decision

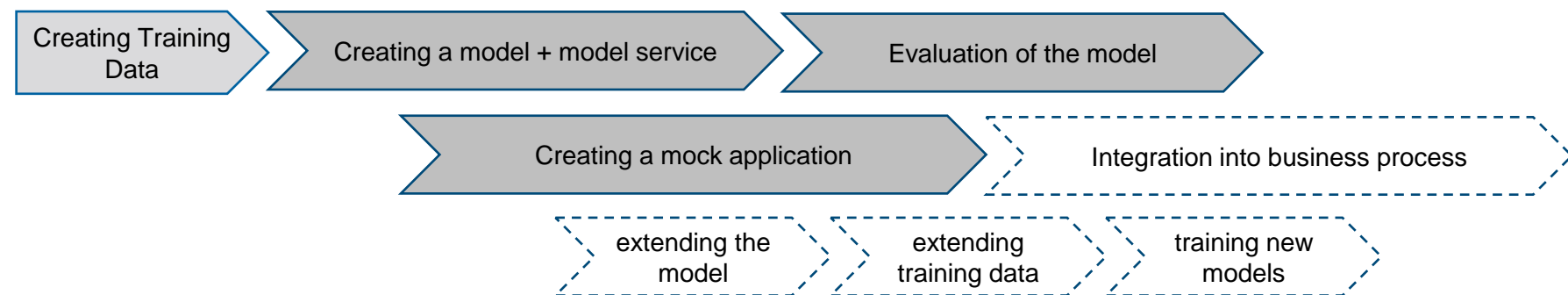


- In ML terms: Multiclass/multilabel Classification Task (≈ 20 categories)

Research Project Automatic Annotation

Starting Point

- At the start of the project (early 2022, pre „GPT breakthrough“):
 - No German-language domain specific (i.e. financial) language model available
 - Hence 2-Step modelling process:
 - (1) Fine-tuning a language model for German financial documents
 - (2) Training a multilabel classifier on top of the language model
 - No public dataset available -> creating training data is the first step



Creating Training Data



Data collection

a) eine von Netze listed on the Official List of the Luxembourg Stock Exchange and traded on the regulated Market "Bourse de Luxembourg" or publicly offered in the Grand Duchy of Luxembourg, the initial Terms will be published in electronic form on the website of the Luxembourg Stock Exchange (www.bourse.lu). Furthermore, the aforementioned Final Terms will be published in a static form on the website of EIBank AG (www.eibank.lu) and on the website of any other stock exchange or trading platform on any regulated market or publicly offered in one or more member states of the European Economic Area (excluding the website of the Final Terms, which will be published in electronic form on the website of EIBank AG (www.eibank.lu)).

15 February 2022
 15 February 2022

Endgültige Bedingungen
Final Terms

EUR 10.000.000 Mann Celsius Future (Interest Swap) Series of 2022/2027 (the "Notes")
 EUR 10.000.000 mannfach von Celsius Future (Interest Swap) Series of 2022/2027 (die "Noten")
 Schuldverschreibungen von 2022/2027 (die "Schuldverschreibungen")

issued and managed by
 EIBank AG
 basierend aufgrund des

EIBANK AG
Debit Issuance Programme
 dated 4 June 2021
 dated 4 June 2021

of
 EIBank AG

Deutsche Zentral-Genossenschaftsbank, Frankfurt am Main
 (having its registered office at: EIBank AG, 60325 Frankfurt am Main, Federal Republic of Germany)

(mit eingetragenem Sitz in Platz der Republik, 60325 Frankfurt am Main, Bundesrepublik Deutschland)

Issued Pursuant to the
 Issuance of 100 per cent during the subscription period
 from 15 February 2022 to 15 February 2022 (in each case including)
 The selling price of the Notes is to be fixed after the expiry period

Ausgegeben: 100 % während der Zeichnungsfrist
 vom 15. Februar 2022 bis 15. Februar 2022 (in jedem Falle einschließlich)
 Nach Ablauf der Zeichnungsfrist ist der Verkaufspreis der Schuldverschreibungen festzulegen

Issued Date: 17 February 2022
 Vervielfältigungsfrist: 17. Februar 2022
 Series Nr.: A1702
 Series Nr.: A1702

Number of prospectus:	413
Issuing period:	2021 - 2022
Eligible documents:	369
Ineligible documents:	44
Training set:	272
Test set:	141 + 141

Data annotation

PART I: TERMS AND CONDITIONS
TEIL I: ANLEHNERBEDINGUNGEN

The PART I of these Final Terms ("Anlehnbedingungen") covers the A1 Terms and Conditions of Fixed Rate Preferred Senior Notes (the "Terms and Conditions") set forth in the Prospectus, and the Additional Terms and Conditions defined in the PART II of these Final Terms, and have the same meanings specified in the Terms and Conditions.

Der Teil I dieser Endklauseln ("Anlehnbedingungen") umfasst die A1-Anlehnbedingungen für festverzinsliche, bevorzugte Senior-Schuldenscheine (die "Anlehnbedingungen"), die im Prospektus und in den zusätzlichen Endklauseln definiert sind, die in dem Teil II dieser Endklauseln festgelegt sind, und haben dieselbe Bedeutung, wie in den Anlehnbedingungen spezifiziert.

All references in this PART I to these Final Terms is numbered paragraphs and sub-paragraphs in the Terms and Conditions, and the Additional Terms and Conditions.

Bezugnahmen in diesem Teil I dieser Endklauseln auf Paragraphen und Subparagraphen beziehen sich auf die Paragraphen und Absätze der Anlehnbedingungen.

References in this PART I to these Final Terms, the Terms and Conditions, taken together, and constitutes the terms and conditions applicable to the Tranche of the Notes, the "Conditions".

Bezugnahmen in diesem Teil I dieser Endklauseln auf Paragraphen und Subparagraphen zusammengefasst, bilden die Bedingungen, die für die Tranche von Schuldverschreibungen anwendbar sind. Bedingungen der Anlehnbedingungen, der zusätzlichen Endklauseln.

Language of Conditions
Sprache der Bedingungen

German and English (German text (controlling and binding))
 Deutsch und Englisch (Deutsch-Text (maßgebend und verbindend))

CURRENT / DENOMINATION FORM / DEFINITIONS
WAHRUNG / STELLUNG / FORM / DEFINITIONEN

- Sub-paragraph (1)
- Sub-paragraph (2)
- Sub-paragraph (3)
- Sub-paragraph (4)
- Sub-paragraph (5)
- Sub-paragraph (6)
- Sub-paragraph (7)
- Sub-paragraph (8)
- Sub-paragraph (9)
- Sub-paragraph (10)
- Sub-paragraph (11)
- Sub-paragraph (12)
- Sub-paragraph (13)
- Sub-paragraph (14)
- Sub-paragraph (15)
- Sub-paragraph (16)
- Sub-paragraph (17)
- Sub-paragraph (18)
- Sub-paragraph (19)
- Sub-paragraph (20)
- Sub-paragraph (21)
- Sub-paragraph (22)
- Sub-paragraph (23)
- Sub-paragraph (24)
- Sub-paragraph (25)
- Sub-paragraph (26)
- Sub-paragraph (27)
- Sub-paragraph (28)
- Sub-paragraph (29)
- Sub-paragraph (30)
- Sub-paragraph (31)
- Sub-paragraph (32)
- Sub-paragraph (33)
- Sub-paragraph (34)
- Sub-paragraph (35)
- Sub-paragraph (36)
- Sub-paragraph (37)
- Sub-paragraph (38)
- Sub-paragraph (39)
- Sub-paragraph (40)
- Sub-paragraph (41)
- Sub-paragraph (42)
- Sub-paragraph (43)
- Sub-paragraph (44)
- Sub-paragraph (45)
- Sub-paragraph (46)
- Sub-paragraph (47)
- Sub-paragraph (48)
- Sub-paragraph (49)
- Sub-paragraph (50)
- Sub-paragraph (51)
- Sub-paragraph (52)
- Sub-paragraph (53)
- Sub-paragraph (54)
- Sub-paragraph (55)
- Sub-paragraph (56)
- Sub-paragraph (57)
- Sub-paragraph (58)
- Sub-paragraph (59)
- Sub-paragraph (60)
- Sub-paragraph (61)
- Sub-paragraph (62)
- Sub-paragraph (63)
- Sub-paragraph (64)
- Sub-paragraph (65)
- Sub-paragraph (66)
- Sub-paragraph (67)
- Sub-paragraph (68)
- Sub-paragraph (69)
- Sub-paragraph (70)
- Sub-paragraph (71)
- Sub-paragraph (72)
- Sub-paragraph (73)
- Sub-paragraph (74)
- Sub-paragraph (75)
- Sub-paragraph (76)
- Sub-paragraph (77)
- Sub-paragraph (78)
- Sub-paragraph (79)
- Sub-paragraph (80)
- Sub-paragraph (81)
- Sub-paragraph (82)
- Sub-paragraph (83)
- Sub-paragraph (84)
- Sub-paragraph (85)
- Sub-paragraph (86)
- Sub-paragraph (87)
- Sub-paragraph (88)
- Sub-paragraph (89)
- Sub-paragraph (90)
- Sub-paragraph (91)
- Sub-paragraph (92)
- Sub-paragraph (93)
- Sub-paragraph (94)
- Sub-paragraph (95)
- Sub-paragraph (96)
- Sub-paragraph (97)
- Sub-paragraph (98)
- Sub-paragraph (99)
- Sub-paragraph (100)
- Sub-paragraph (101)
- Sub-paragraph (102)
- Sub-paragraph (103)
- Sub-paragraph (104)
- Sub-paragraph (105)
- Sub-paragraph (106)
- Sub-paragraph (107)
- Sub-paragraph (108)
- Sub-paragraph (109)
- Sub-paragraph (110)
- Sub-paragraph (111)
- Sub-paragraph (112)
- Sub-paragraph (113)
- Sub-paragraph (114)
- Sub-paragraph (115)
- Sub-paragraph (116)
- Sub-paragraph (117)
- Sub-paragraph (118)
- Sub-paragraph (119)
- Sub-paragraph (120)
- Sub-paragraph (121)
- Sub-paragraph (122)
- Sub-paragraph (123)
- Sub-paragraph (124)
- Sub-paragraph (125)
- Sub-paragraph (126)
- Sub-paragraph (127)
- Sub-paragraph (128)
- Sub-paragraph (129)
- Sub-paragraph (130)
- Sub-paragraph (131)
- Sub-paragraph (132)
- Sub-paragraph (133)
- Sub-paragraph (134)
- Sub-paragraph (135)
- Sub-paragraph (136)
- Sub-paragraph (137)
- Sub-paragraph (138)
- Sub-paragraph (139)
- Sub-paragraph (140)
- Sub-paragraph (141)
- Sub-paragraph (142)
- Sub-paragraph (143)
- Sub-paragraph (144)
- Sub-paragraph (145)
- Sub-paragraph (146)
- Sub-paragraph (147)
- Sub-paragraph (148)
- Sub-paragraph (149)
- Sub-paragraph (150)
- Sub-paragraph (151)
- Sub-paragraph (152)
- Sub-paragraph (153)
- Sub-paragraph (154)
- Sub-paragraph (155)
- Sub-paragraph (156)
- Sub-paragraph (157)
- Sub-paragraph (158)
- Sub-paragraph (159)
- Sub-paragraph (160)
- Sub-paragraph (161)
- Sub-paragraph (162)
- Sub-paragraph (163)
- Sub-paragraph (164)
- Sub-paragraph (165)
- Sub-paragraph (166)
- Sub-paragraph (167)
- Sub-paragraph (168)
- Sub-paragraph (169)
- Sub-paragraph (170)
- Sub-paragraph (171)
- Sub-paragraph (172)
- Sub-paragraph (173)
- Sub-paragraph (174)
- Sub-paragraph (175)
- Sub-paragraph (176)
- Sub-paragraph (177)
- Sub-paragraph (178)
- Sub-paragraph (179)
- Sub-paragraph (180)
- Sub-paragraph (181)
- Sub-paragraph (182)
- Sub-paragraph (183)
- Sub-paragraph (184)
- Sub-paragraph (185)
- Sub-paragraph (186)
- Sub-paragraph (187)
- Sub-paragraph (188)
- Sub-paragraph (189)
- Sub-paragraph (190)
- Sub-paragraph (191)
- Sub-paragraph (192)
- Sub-paragraph (193)
- Sub-paragraph (194)
- Sub-paragraph (195)
- Sub-paragraph (196)
- Sub-paragraph (197)
- Sub-paragraph (198)
- Sub-paragraph (199)
- Sub-paragraph (200)
- Sub-paragraph (201)
- Sub-paragraph (202)
- Sub-paragraph (203)
- Sub-paragraph (204)
- Sub-paragraph (205)
- Sub-paragraph (206)
- Sub-paragraph (207)
- Sub-paragraph (208)
- Sub-paragraph (209)
- Sub-paragraph (210)
- Sub-paragraph (211)
- Sub-paragraph (212)
- Sub-paragraph (213)
- Sub-paragraph (214)
- Sub-paragraph (215)
- Sub-paragraph (216)
- Sub-paragraph (217)
- Sub-paragraph (218)
- Sub-paragraph (219)
- Sub-paragraph (220)
- Sub-paragraph (221)
- Sub-paragraph (222)
- Sub-paragraph (223)
- Sub-paragraph (224)
- Sub-paragraph (225)
- Sub-paragraph (226)
- Sub-paragraph (227)
- Sub-paragraph (228)
- Sub-paragraph (229)
- Sub-paragraph (230)
- Sub-paragraph (231)
- Sub-paragraph (232)
- Sub-paragraph (233)
- Sub-paragraph (234)
- Sub-paragraph (235)
- Sub-paragraph (236)
- Sub-paragraph (237)
- Sub-paragraph (238)
- Sub-paragraph (239)
- Sub-paragraph (240)
- Sub-paragraph (241)
- Sub-paragraph (242)
- Sub-paragraph (243)
- Sub-paragraph (244)
- Sub-paragraph (245)
- Sub-paragraph (246)
- Sub-paragraph (247)
- Sub-paragraph (248)
- Sub-paragraph (249)
- Sub-paragraph (250)
- Sub-paragraph (251)
- Sub-paragraph (252)
- Sub-paragraph (253)
- Sub-paragraph (254)
- Sub-paragraph (255)
- Sub-paragraph (256)
- Sub-paragraph (257)
- Sub-paragraph (258)
- Sub-paragraph (259)
- Sub-paragraph (260)
- Sub-paragraph (261)
- Sub-paragraph (262)
- Sub-paragraph (263)
- Sub-paragraph (264)
- Sub-paragraph (265)
- Sub-paragraph (266)
- Sub-paragraph (267)
- Sub-paragraph (268)
- Sub-paragraph (269)
- Sub-paragraph (270)
- Sub-paragraph (271)
- Sub-paragraph (272)
- Sub-paragraph (273)
- Sub-paragraph (274)
- Sub-paragraph (275)
- Sub-paragraph (276)
- Sub-paragraph (277)
- Sub-paragraph (278)
- Sub-paragraph (279)
- Sub-paragraph (280)
- Sub-paragraph (281)
- Sub-paragraph (282)
- Sub-paragraph (283)
- Sub-paragraph (284)
- Sub-paragraph (285)
- Sub-paragraph (286)</

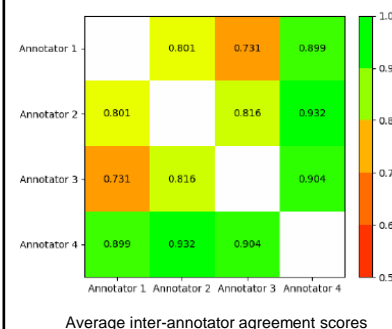
Annotation tool: Konfuzio
Annotation types: ~40

Disregarding of pages without annotations for training and validation purposes.

Annotation statistics

Target type	Train	Test
coupon fixed	431	375
coupon variable index	56	84
coupon variable margin	38	42
coupon variable operator	37	43
coupon variable tenor	45	75
currency	514	577
early redemption amount	64	52
early redemption	177	108
isin	421	417
principal amount	784	800
redemption at maturity amount	26	42
redemption at maturity	370	347
special termination	96	109
special termination amount	61	63
status non preferred	56	47
status senior non preferred	488	333
type of instrument	431	422

Inter-annotator agreement

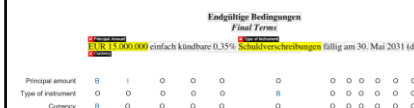


Test set was used to measure IAA. Therefore, every prospectus in the test set was annotated by a second analyst. 4 analysts served as annotators in total.

Data preprocessing

1st step: Extraction of JSON-formatted raw data containing the annotations from the annotation tool

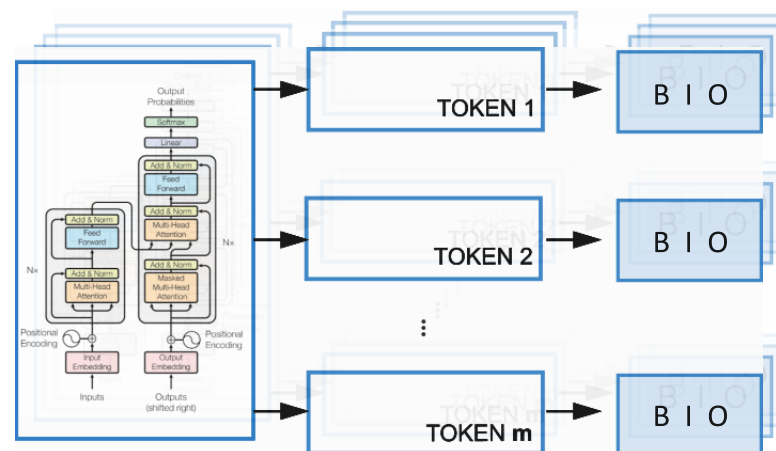
2nd step: Conversion and transformation of extracted data into dataset for token classification (BIO encoding)



- Implementation of dataset classes using Hugging Face Datasets framework
- Challenge: overlapping text sequences belonging to different annotation types

1. Conversion PDF -> Text (including OCR)
2. Text processing and clean-up (e.g. extraction of German parts of bilingual docs, analysis of textboxes, ...)
3. *Embedding (Text to vectors) using fine tuned language model*
4. *Labelling with multilabel classifier*
5. Decision based on deterministic rules (derived from EU Guideline)

3 + 4

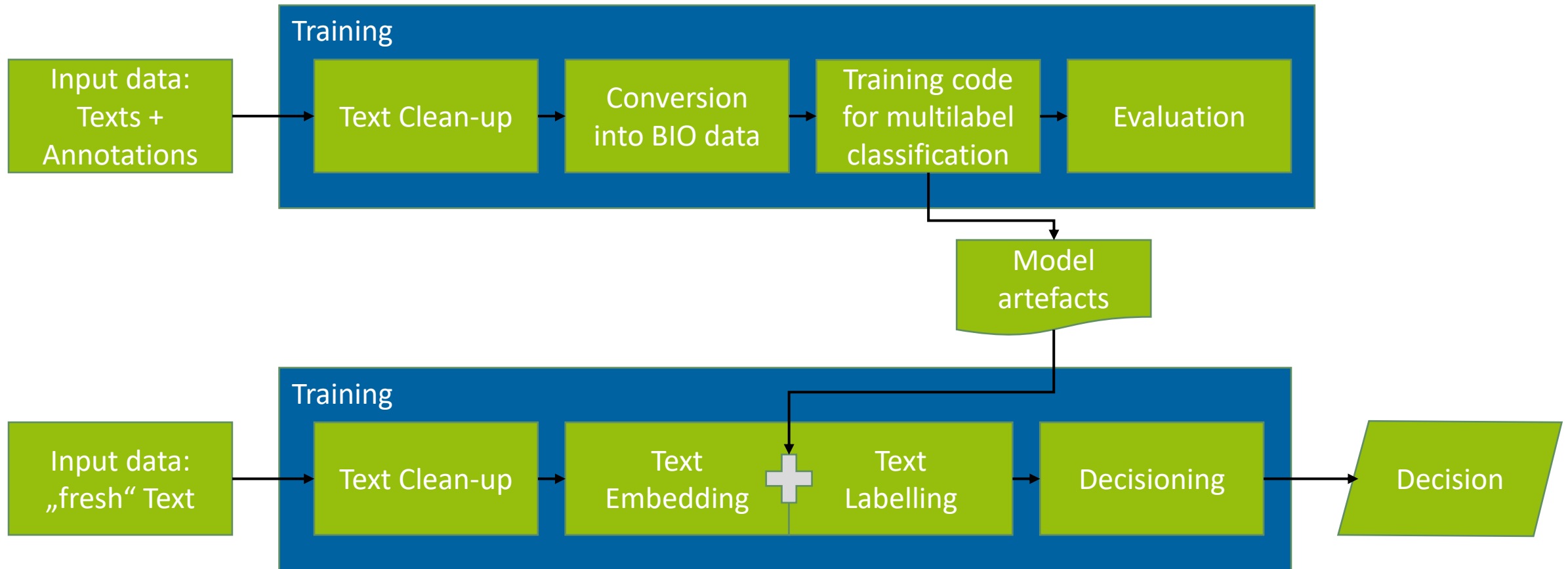


Integration

of the model into the business process

Operating the Model

Model Training and Decisioning



Integration into Business Process

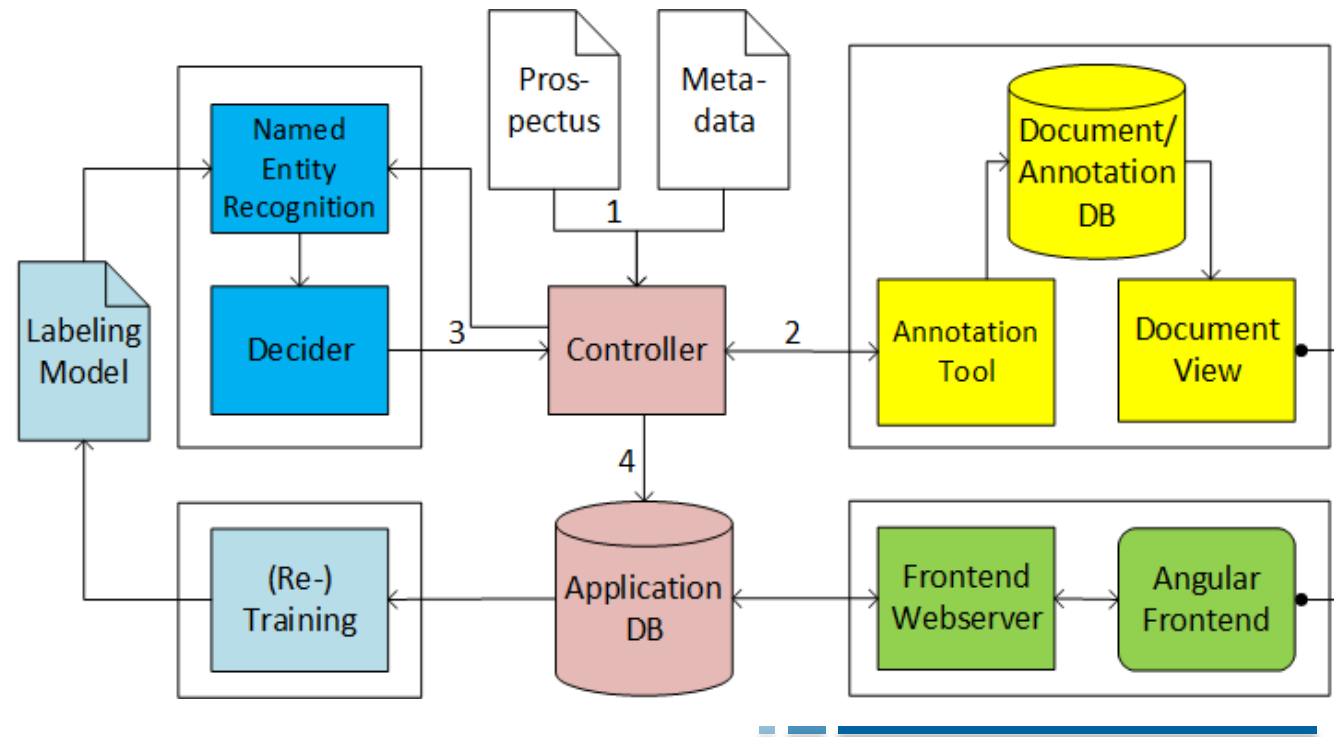
Process Needs

- Given a document the experts needs:
 - a. the decision of the model,
 - b. the criteria causing that decision and
 - c. (optimally) the relevant passages in the document (or relevant meta data) to
- check the validity of the ML decision.
- If the model makes a mistake, the expert needs to **overwrite that decision** and
- (optimally) collect the data for future model improvements (retraining)
- If retraining is undertaken, we need both valid as well as invalid model decisions.

Integration into Business Process

Overview of Application Architecture

- Containerized application with communication via REST
- Integration into the actual business application (MARS) open as of yet



Implications of Using ML in the Process



- Using an ML model can reduce processing time by replacing manual reading with reviewing found passages



- An ML model will always have a chance for error, but the **accuracy can reach the Inter Annotator Agreement (IAA)** at the least



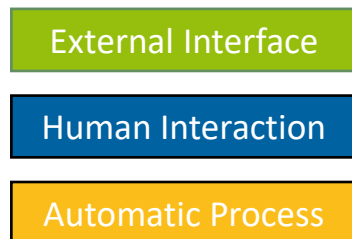
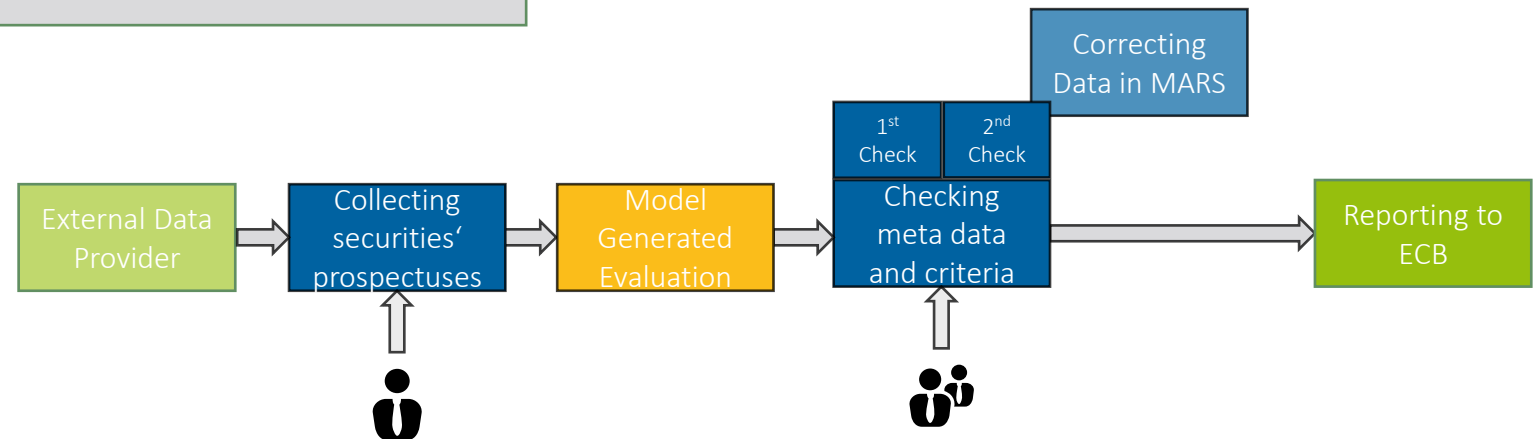
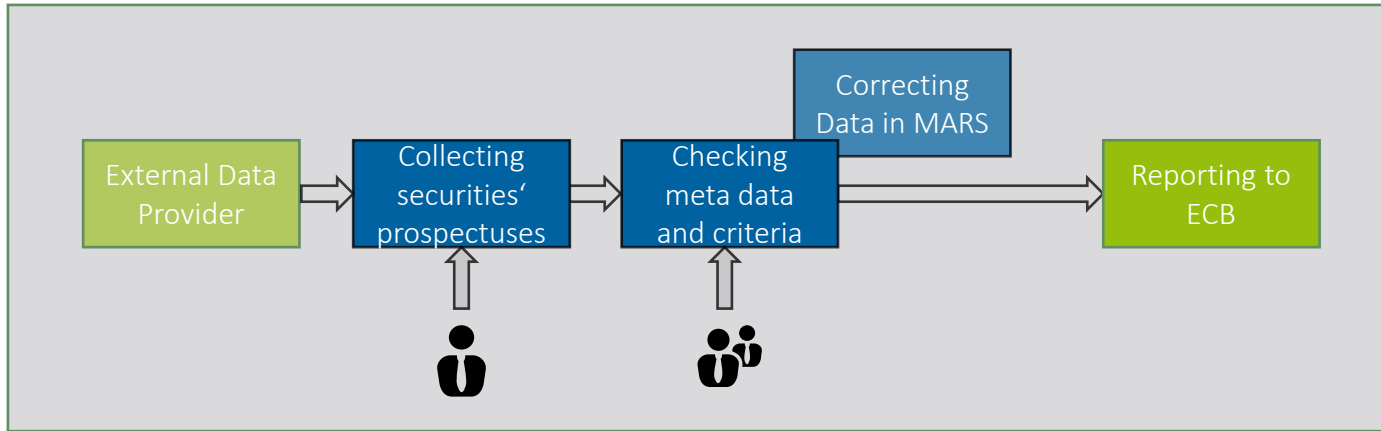
- Current legal environment requires a „**human in the loop**“
 - If model accuracy is (acceptably) high, the four eyes principle (as well as the review by two experts) could be replaced by a simple review
(2 pairs of human eyes \Rightarrow „AI eyes“ + 1 pair of human eyes)



- Using an ML model will require:
 - **continuous monitoring** of model performance
 - **continuous improvement** of model mistakes and training data

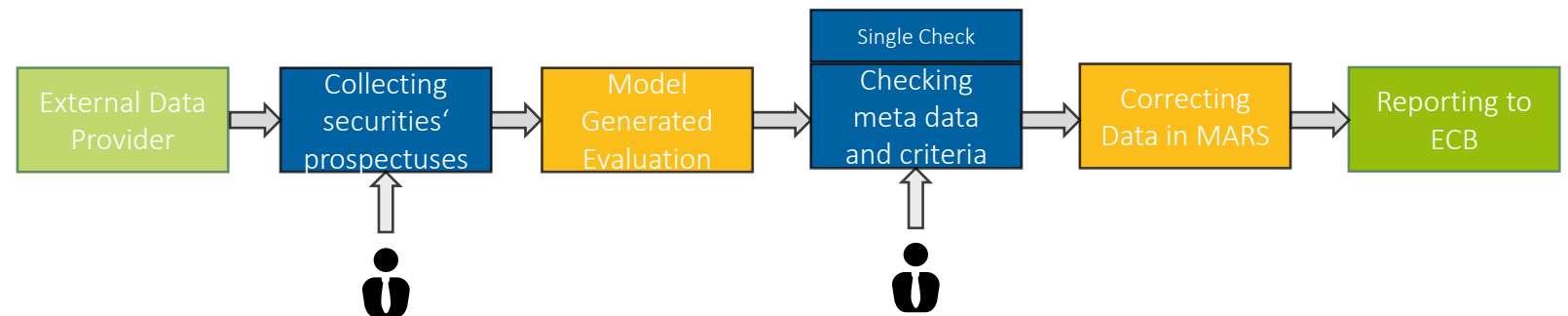
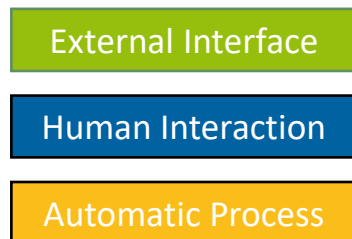
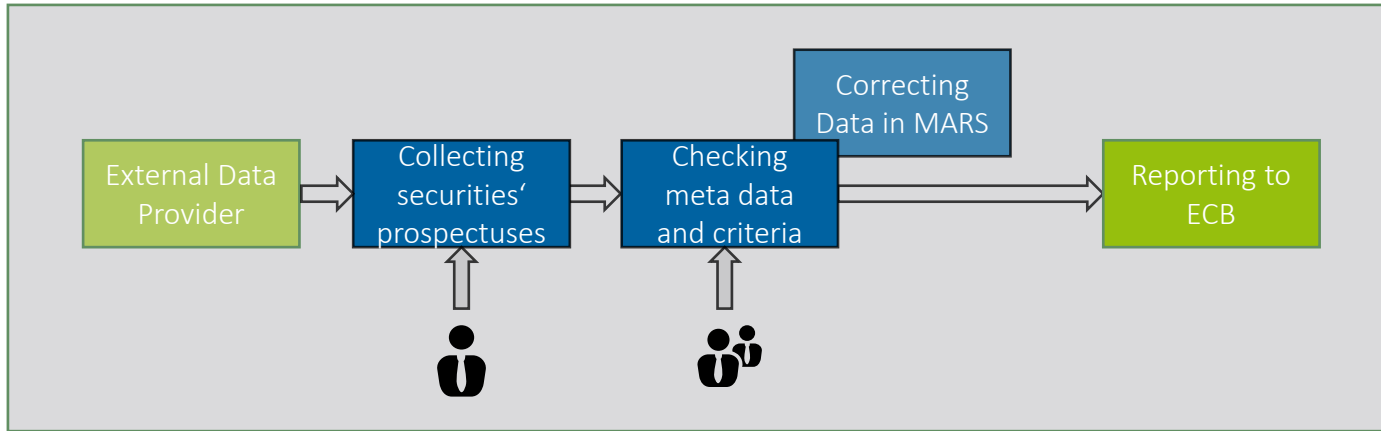
Evolving the „4 Eyes Principle“

New Process Flow – Proof of Concept



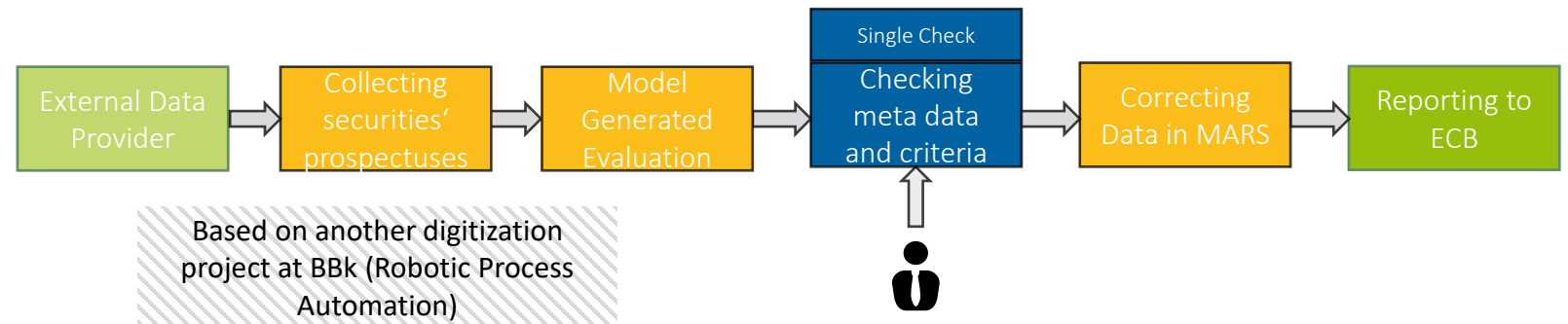
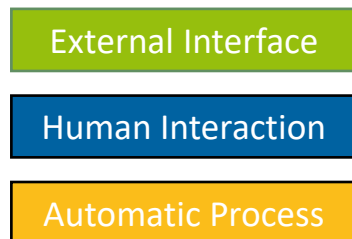
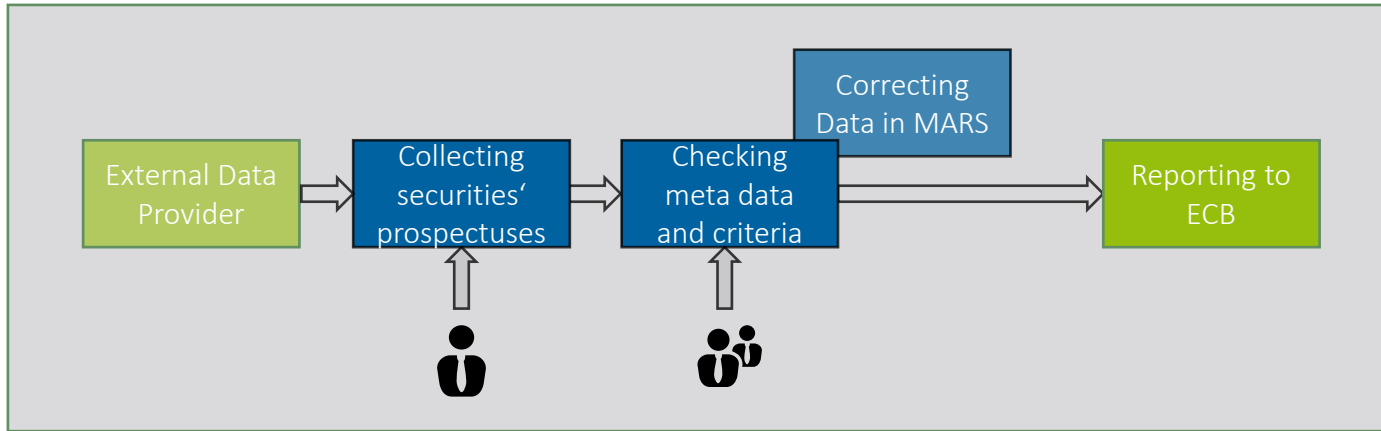
Evolving the „4 Eyes Principle“

New Process Flow – 1st Evolution



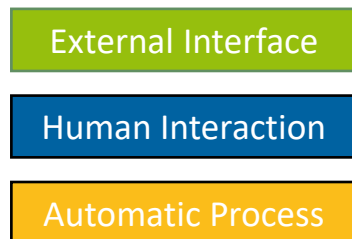
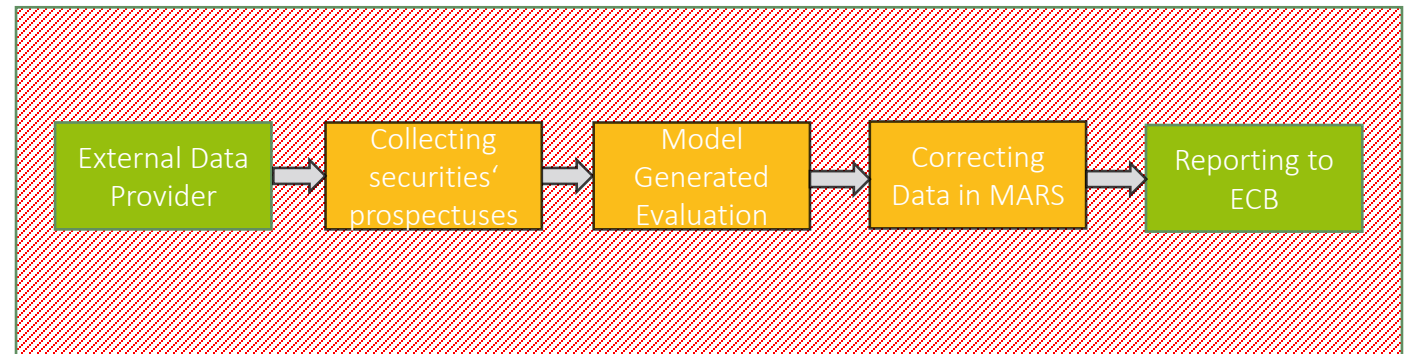
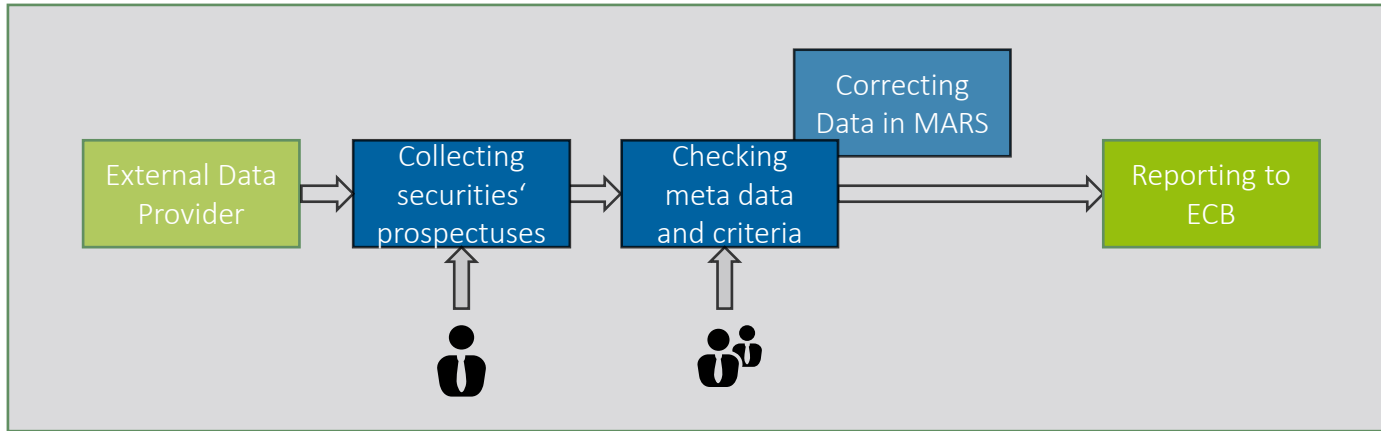
Evolving the „4 Eyes Principle“

New Process Flow – 2nd Evolution



Evolving the „4 Eyes Principle“

Currently not Possible: Fully Automated Process – No Human in the Loop



Learnings

from the Process

Learnings from the Project



- Creating training data is highly costly



- Understanding the business process is key
 - if only part of the process is automated, **the benefit may not outweigh the complexity**



- Building the necessary environment is highly complex
 - The codebase of the proof of concept easily reaches **10'000 lines of code**



- **Integration into production is hard**, in particular if it necessitates new components, e.g.
 - Application for creating and storing text annotations
 - ML model monitoring and model archives (MLOps)
 - GPUs for model training

Questions?

SCAI

Service and Community Center for Artificial Intelligence

scai@bundesbank.de