



**TMF** WEEK  
2022

by **montrium**

# Can Machine Learning Provide TMF Benefits beyond Processing Documents

Paul Fenton  
Chief Executive Officer @ Montrium  
&  
Michael Zwetkow  
Chief Innovation Officer @ Montrium

# Agenda

- What is Artificial Intelligence
- Where can it be applied in TMF today and in the future
- AI-Assisted QC and Cross-Checks
- Risk based approach
- Predictive completeness
- The importance of human feedback

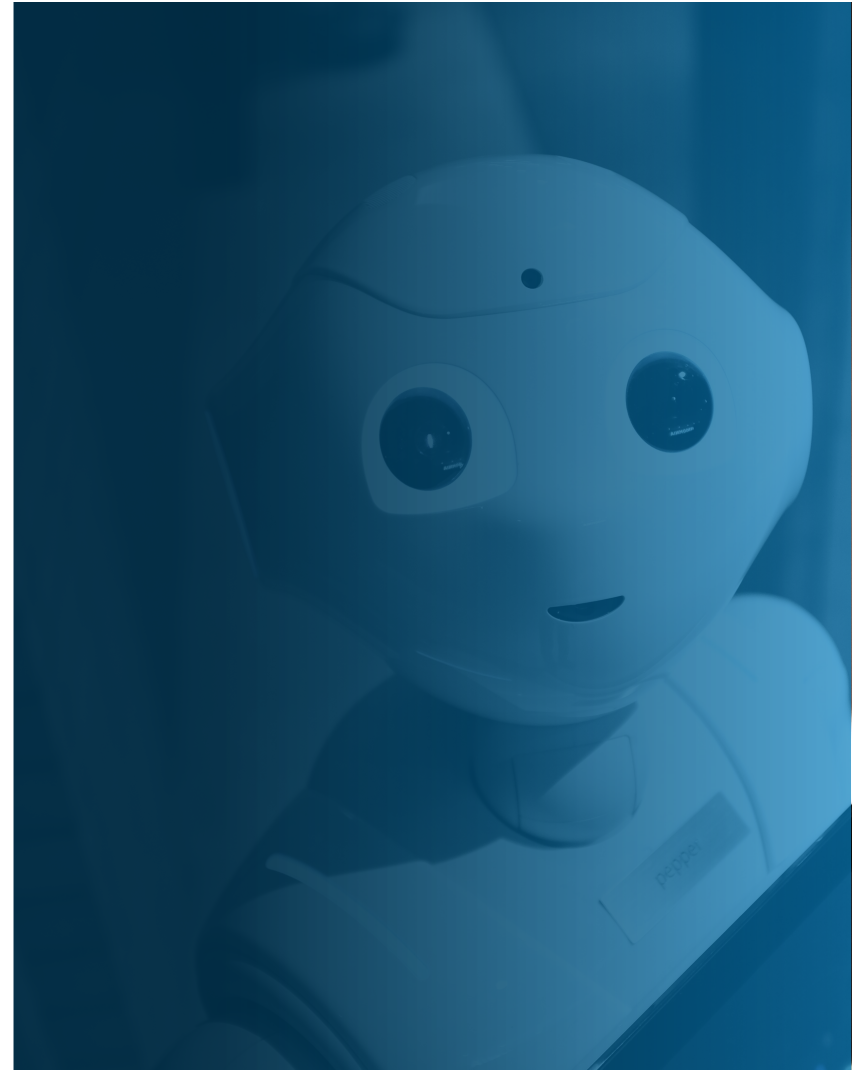


# What is Artificial Intelligence?

montrium

noun: **artificial intelligence**;  
the theory and development of  
computer systems able  
to perform tasks that normally  
require human intelligence,  
such as visual perception, speech  
recognition, decision-making, and  
translation between languages.

- The Oxford Dictionary



# Examples of AI in our everyday lives

- Face ID to unlock your phone
- Google search results
- Alexa....What is the weather forecast today?
- Find the best route with Waze
- Netflix, show me something I would like to watch
- Autocomplete in emails and texts (not always so intelligent 😊)
- Self-driving cars
- And many more....



# How does AI learn?

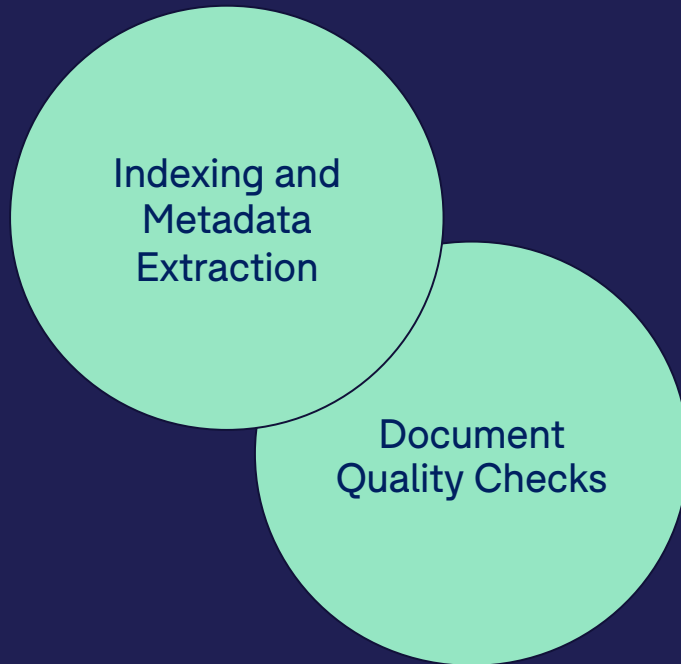
- AI consists of iterative, 'learning' algorithms that utilize:
  - **(Big) data + high computing power** → to make interpretations, predictions or decisions in an autonomous or semi-autonomous fashion
  - **Machine learning** → derive rules or patterns without the latter being explicitly programmed (supervised vs unsupervised)
  - **Deep learning** → subset of machine learning where algorithms transform data over a series of steps (aka layers)

A dark blue background with a faint image of a human hand on the left and a white robotic hand on the right, both reaching towards the center. The text is overlaid in the center.

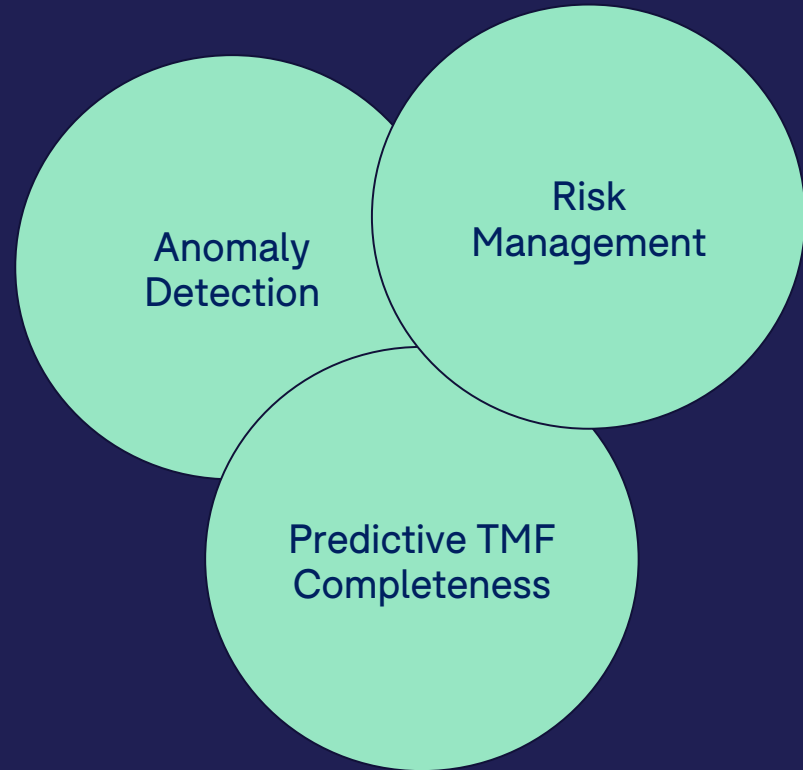
So how can we leverage AI and ML in  
eTMF?

montrium

## Today



## Tomorrow



montrium

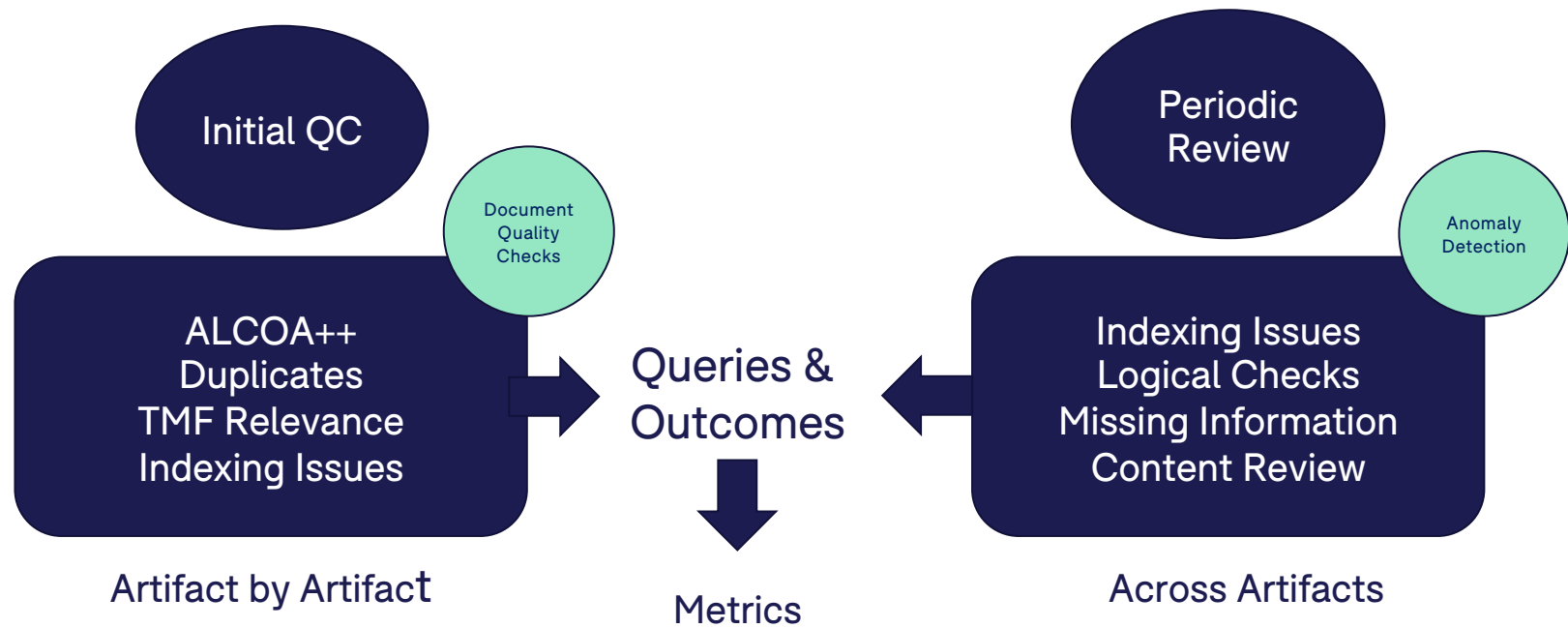


# What is TMF Quality?

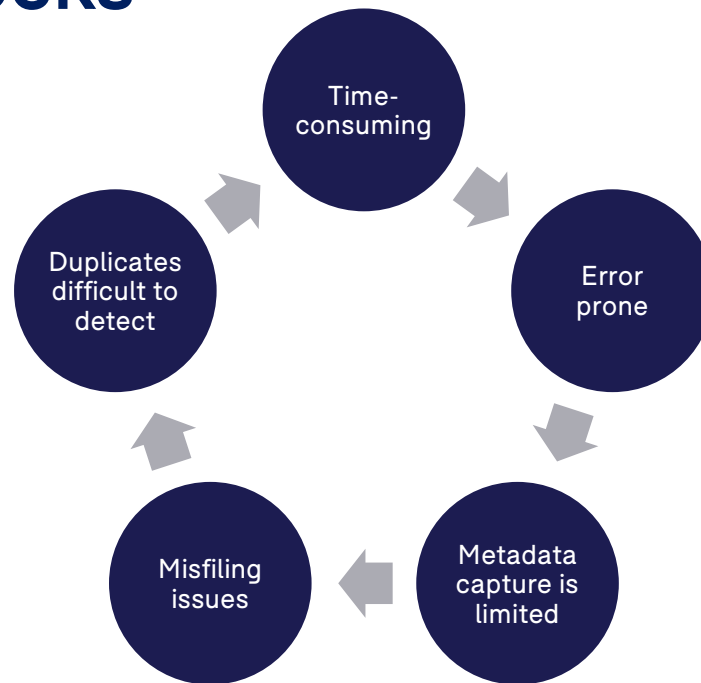
- All artifacts:
  - Are TMF relevant
  - Meet the requirements of ALCOA++
  - Correctly indexed and filed
  - Are not duplicated
  - Adequately versioned
  - Readily retrievable
  - Tell a complete and accurate story of what happened in the study
  - Have a complete audit trail and activity log available for inspection



# How do we ensure quality?



# Challenges with QC and Cross-Checks







**How can Artificial Intelligence help?**

montrium

# AI / ML-driven capabilities



Revealing meaningful insights from large volumes of documents and data

## Document indexing

- Bulk migration
- Addition of individual / small numbers of documents (e.g. drag and drop)
- Category alternative suggestions
- Template matching

## Quality checks - individual docs

- Duplicate documents
- Missing signatures
- Missing date
- Document translations
- Document legibility
- Missing pages
- Incorrectly named documents
- Accurate headers and footers
- Correct orientation for viewing
- PDF is formatted correctly
- Identify draft versions
- Improper corrections
- Identify personal data that shouldn't be there


## Document relationship identification

- Metadata identification
- Document connectedness
- Document clustering

# Visual Example: Metadata extraction

Metadata extraction works by identifying key pieces of information within a document, including:

- Key value pairs
- Names of people
- Email addresses
- Dates
- Locations (Country, State, City)

  
 STATE OF TENNESSEE  
 DEPARTMENT OF HEALTH  
 DIVISION OF HEALTH LICENSURE AND REGULATION  
 DIVISION OF HEALTH RELATED BOARDS  
 665 Mainstream Dr.  
 Nashville, TN 37243  
[tn.gov/health](http://tn.gov/health)  
 TENNESSEE BOARD OF MEDICAL EXAMINERS  
 1-800-778-4123 or (615) 532-4384

**November 7, 2014**

TO WHOM IT MAY CONCERN:

This verification can be considered primary source. To expedite the verification process, this is the standard format used by the Board of Medical Examiners. The Board of Medical Examiners is pleased to furnish the following information from our files:

<b>PROFESSION</b>	Medical Doctor
<b>RANK</b>	Medical Doctor
<b>NAME</b>	John Doe
<b>LICENSE NUMBER</b>	1234
<b>ISSUE DATE</b>	09-03-2011
<b>EXPIRATION DATE</b>	07-01-2024
<b>STATUS</b>	Licensed
<b>STATUS DATE</b>	09-14-2011
<b>SPECIAL ENDORSEMENTS</b>	Internal Medicine Medical Oncology

**COMMENTS:** There is no history of disciplinary action on this license. The State of Tennessee only provides the above information. Any other information needed must be obtained from the licensee.

Sincerely,  
 Tennessee Board of Medical Examiners

■ person  
■ key value pair  
■ date  
■ custom info  
■ page number  
■ email  
■ signatures



# Benefits of AI-assisted metadata extraction

- We can extract a lot more data! Not just for filing, but for telling a better story...
- Artifacts can be cross-referenced more easily to multiple individuals, sites, or countries, i.e. Central IRB approvals, 1572s etc.
- Reduces manual data entry significantly
- Reduces data entry errors
- Significantly improves our ability to perform cross checks

# Challenges with AI-assisted metadata extraction

- Can result in misfiling if values are incorrectly identified
- There could be competing values for the same metadata field, i.e. approval date appears several times in the document
- Managing data formats especially around date values



# Examples of AI-assisted quality checks

## Missing information:

- Signatures
- Dates
- Pages

Pharma Co

ABC123 Site Request for Study Drug Kit

Legend

- Person
- Country
- Key site user
- Date
- Custom info
- Page number
- Email
- Signature

Study Acronym: ABC123

Site Number: 0001

Site Name: Acme University

Site Address: 1234 Central Ave, Suite 1000, Cambridge, MA, 1000

Requested By:

Name: Jane Doe

Signature: [Redacted] ▲ Unsigned signature field

Date: May 1, 2018

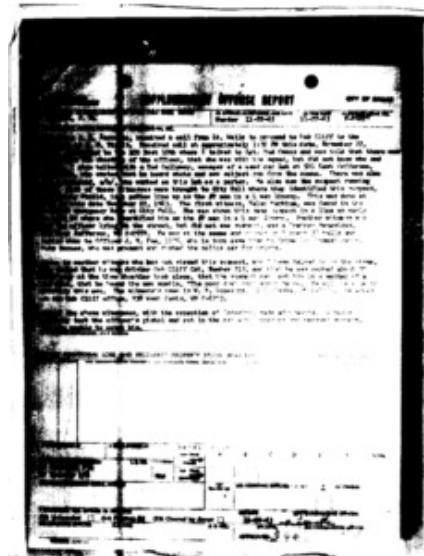
Comments:

Instructions: Email a copy of this form to [john.smith@pharmaco.com](mailto:john.smith@pharmaco.com) and [betty.white@acme.edu](mailto:betty.white@acme.edu)

Page 1/1

## Poorly scanned documents:

- Low resolution
- Obscured text
- Incorrect orientation



## Process or policy violations:

- Filling of duplicate documents
- Draft documents marked as "final"
- Documents containing PII but should not

Protocol Title: Protocol -2022-LTD- Version: 1.0- 19 April 2022

1.3 SCHEDULE OF ACTIVITIES (SOA)

The schedule of activities captures the procedures that will be accomplished at each study visit, and all contact, with study participants e.g., telephone contacts. This includes any tests that are used for eligibility, participant randomization or stratification, or decisions on study intervention discontinuation. Only include procedures that contribute to participant eligibility and study objectives and endpoints. Other procedures should be done sparingly and with consideration, as they may add unnecessary complexity and distract from recruitment.

	Screening Visit 1	Baseline/Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18	Week 19	Week 20	
Procedures																						
Eligibility screen	X																					
Demographics	X																					
Medical history	X																					
Randomization	X																					
Administer study intervention	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant medication review	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Physical exam (including height and weight)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ECG	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Weight	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Vitals	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Performance status	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Neurologic	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Speech/chemistry	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Psychiatry	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ECG (as indicated)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Adverse event review and evaluation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Biopsy/Imaging (as indicated)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Other assessments (e.g., knowledge assay, consent/assent)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant Case Report Forms (CRFs)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Health care provider (HCP) visits (including, but not limited to, baseline, baseline, BLK, calcium, cholest, creatinine, glucose, LDK, phosphorus, potassium, total protein, AST, ALT, sodium)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Serial pregnancy test (over the course of the study)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Michael Zwetkoff

To be confirmed

Tip: Press Ctrl+Enter to post.

# Benefits of AI-assisted QC

- We can do a lot more verifications with little or no human intervention
- Queries should be generated automatically or even in real-time if documents that are being uploaded do not pass the tests
- The TMF management team can focus more on cross checks and periodic reviews rather than on basic QC checks
- These checks could also be run in batch when migrating content into the TMF

montrium

# Benefits of AI-assisted oversight



Average Document Type Prediction Confidence Levels Across Studies

DocumentType (actual)	Study 1	Study 2	Study 3	Total
FINANCIAL DISCLOSURE FORM	1.00	0.99	1.00	1.00
IP RE-LABELING DOCUMENTATION	1.00			1.00
DATABASE LOCK MEMO		0.99		0.99
INSURANCE CERTIFICATE			0.99	0.99
ICH-GCP TRAINING CERTIFICATE	0.99	0.99	0.99	0.99
THIRD PARTIES CONFIDENTIALITY AGREEMENT	0.99			0.99
SUBJECT QUESTIONNAIRE	0.98	1.00		0.98
SOURCE DATA	0.98			0.98
MEDICAL LICENSE		0.99	0.97	0.98
MONITORING VISIT REPORT	0.98	0.98		0.98
MONITORING VISIT CONFIRMATION LETTER	0.99	0.82	0.99	0.98
MONITORING VISIT FOLLOW-UP DOCUMENTATION	0.99	0.87		0.97
NON-IP SHIPMENT DOCUMENTATION	0.97			0.97
OTHER CURRICULUM VITAE	0.99	0.97	0.97	0.97
IP ACKNOWLEDGEMENT OF RECEIPT	0.97	0.98		0.97
TRANSLATION STATEMENT			0.96	0.96
DEBARMENT STATEMENT		0.95	0.95	0.95
CERTIFICATION OR ACCREDITATION		0.95	0.95	0.95
INVESTIGATORS MEETING MATERIAL	0.95			0.95
ACCEPTANCE OF INVESTIGATOR BROCHURE	0.99	0.89	0.99	0.95
ICF REVIEW FORM	0.90	0.99	0.92	0.95
SITE CONFIDENTIALITY AGREEMENT		0.91	0.98	0.95
FDA FORM 1572	0.55	1.00	1.00	0.95
DATA MANAGEMENT PLAN		0.94		0.94
MONITORING VISIT LOG		0.94		0.94
INITIATION VISIT FOLLOW-UP LETTER	0.99	0.95	0.88	0.92
TRIAL TEAM EVIDENCE OF TRAINING	0.96	0.88	0.86	0.91
SUBJECT PARTICIPATION CARD	0.91			0.91
IP STORAGE CONDITION EXCURSION DOCUMENTATION	0.90	0.92	0.94	0.91

ICH GCP Statement of Investigator (non-US sites)

### ICH GCP STATEMENT OF INVESTIGATOR

(See instructions on next page)

1. NAME AND ADDRESS OF INVESTIGATOR

Name of Investigator

Address 1 Address 2

City State/Province/Region Country ZIP or Postal Code

2. EDUCATION, TRAINING, AND EXPERIENCE THAT QUALIFY THE INVESTIGATOR AS AN EXPERT IN THE CLINICAL INVESTIGATION OF THE DRUG FOR THE USE UNDER INVESTIGATION ONE OF THE FOLLOWING IS PROVIDED. Select one of the following:

Curriculum Vitae  Other Statement of Qualifications

3. NAME AND ADDRESS OF INSTITUTION WHERE THE CLINICAL INVESTIGATION WILL BE CONDUCTED

Name of Institution

Address 1 Address 2

City State/Province/Region Country ZIP or Postal Code

4. NAME AND ADDRESS OF ANY CLINICAL LABORATORY FACILITIES TO BE USED IN THE STUDY

5. NAME AND ADDRESS OF THE INDEPENDENT ETHICS COMMITTEE THAT IS RESPONSIBLE FOR REVIEW AND APPROVAL OF THE STUDY

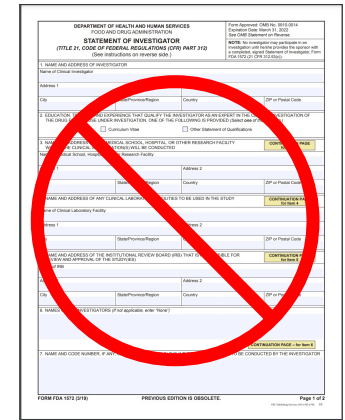
Name of IEC

Address 1 Address 2

City State/Province/Region Country ZIP or Postal Code

6. NAMES OF SUBINVESTIGATORS (if not applicable, enter "None")

FDA FORM 1572



# Cross checks and anomaly management



- Increased metadata means more information to perform cross checks
- AI can use a combination of pre-defined rules and pattern detection to identify anomalies
  - i.e. Ethics submission for protocol amendment and updated CRF does not match approval letter
  - Drug shipment order has a different order number to Drug shipment receipt acknowledgement
  - Site missing DIL when all other sites have one for a specific safety event
- Pattern detection and human feedback can be used by machine learning to improve anomaly detection over time



# Advantages of automated Cross-Checks



- We can systematically perform AI assisted automated cross-checks on all artifacts and as much as we like
- This means that we can significantly increase oversight and improve inspection readiness
- The study team must still review queries and do periodic verification that the AI is still functioning correctly
- A risk-based approach should be used..

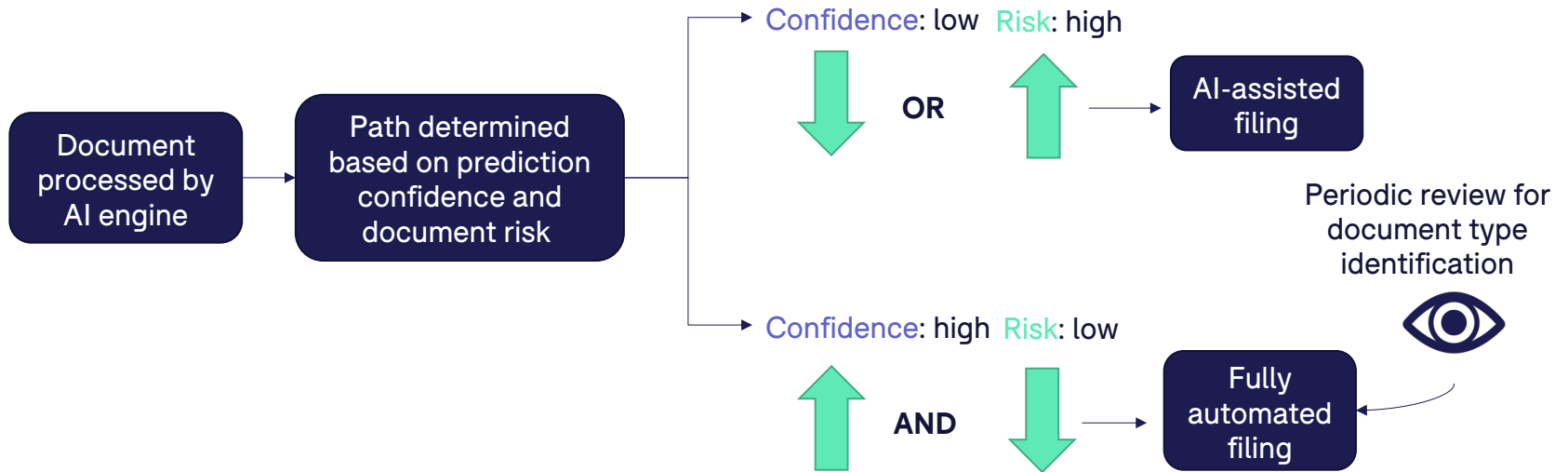
# Predictive completeness

- Today completeness is very challenging as it is dependent on knowledge of events in the clinical trial
- Ideally we should base our TMF on a process model with clear relationships between artifacts
- AI and machine learning could help by
  - Detecting sequences of documents that are related to each other
  - Looking at historical instances of processes and identifying a) which artifacts were collected and b) which associated processes were triggered
- Placeholder generation is still overseen and managed by humans and fed back to machine learning algorithms

## Risk scoring

- Risk based approach to TMF Management is a mandatory requirement
- Risk scoring is complex and is based on many factors
- An ideal solution would be a combination of a risk scoring algorithm which is fed by metadata (extracted by AI) and historical risk trends
- Risk scoring can be used to drive automated cross-checks and other oversight activities

# Balance of confidence and risk



# The importance of human oversight

- AI tools will generally be used to support human activities, not replace them
- The goal is to help the individual focus on activities that provide more value and decrease the risk of human error, by providing suggestions to allow the human to make better informed decisions
- In the context of highly regulated activities or those associated with higher risk, AI tools can assist humans to perform optimally by reducing manual data entry, as well as repetitive tasks