



2026 CDISC China Interchange

Modernizing TMF Management: Applying AI to Drive Quality, Completeness, and Efficiency

Wei Du (杜玮)
Life Sciences Director
TransPerfect



Speakers



Wei Du

Life Sciences Director
TransPerfect

- 13+ years of experience supporting clinical trials in the life sciences industry (including sponsors, CROs, sites, and non-profit organizations), with a focus on language services and electronic clinical solutions.
- Support sponsors and CROs in identifying risks and gaps in TMF management to strengthen inspection readiness and compliance.
- Help clients resolve operational challenges efficiently while reducing GxP-related risks.



5 Most Time-Consuming TMF Tasks?

- Metadata entry
- Document indexing
- Quality checks
- Chasing of missing documents
- Duplication cleanup

5 Most Valuable TMF Tasks?

- Establishment and maintenance of a fit-for-purpose TMF plan
- Risk-based TMF planning and review
- Ensuring inspection readiness and traceability
- Risk identification and trend analysis
- Cross-functional governance and collaboration

5 Most Boring TMF Tasks?

- Metadata entry and maintenance
- Document processing and indexing
- Manual QC of document
- Following up on missing / late document
- Double data entry across systems and reconciliation activities

My Chats
with 豆包,
DeepSeek,
千问,
ChatGPT

...

TMF Automation



Metadata Extraction & Classification



ALCOA++ Document Quality Checks



Mock Inspection and Cross-Doc Edit Checks

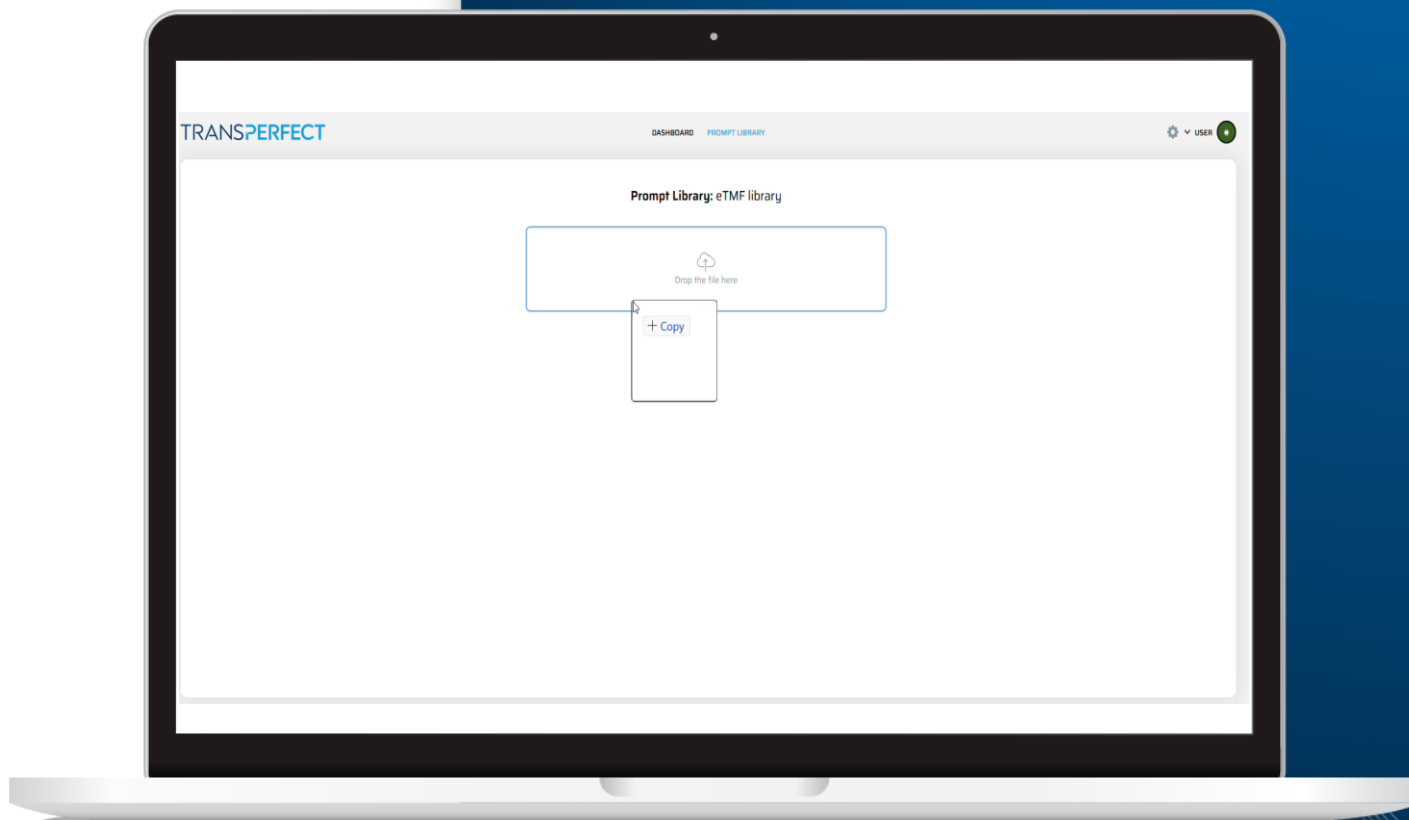


Live TMF and Compliance Copilot



eClinical and eTMF Chatbot

Automation for Metadata Extraction & Classification



- Automates classification & metadata extraction for clinical trial documents
- Pretrained on CDISC model: no heavy training required
- Extracts 25+ metadata fields & flags document issues
- Works in 40+ languages: translation + metadata in one workflow
- Confidence score enables risk-based QC

Automate cut manual TMF coding by up to 70%

Across sponsors and studies, Automate auto-coded 2,032 of 3,439 TMF documents, the work a team would otherwise do by hand.

● Sponsor

70%

decrease in
manual effort

3.4×

productivity
increase

960 of 1,366 documents auto-coded

30 document types fully automated (100%)

● CRO

52%

decrease in
manual effort

2.1×

productivity
increase

1,072 of 2,073 documents auto-coded

58 document types fully automated (100%)

Combined: 59% less manual coding effort across 3,439 documents — 2.4× the throughput of hand-coding.



Automate score = documents automatically coded ÷ total documents, across every document type in each sponsor's TMF. Source: per-room Automate metrics export.

Automation for Document Quality Checks

07_FDA_1571_missing_pages_rotation_probs.pdf

Grid JSON

Quality Score
39

CRITICAL

Summary: This FDA 1571 IND form for Virtual Data Corp / DELPHI Algorithmicin Trialsulfate (proposed indication: treatment of chronic document misclassification disorder and acute regulatory filing anxiety in adult clinical trial populations) has multiple critical quality issues including missing signatures, missing dates, and missing required information fields.

Findings (9)

- Missing Sponsor Signature on Page 3** DQA-381
 Page 3 (labeled 'Page 2 of 3' in footer) shows field 28 'Signature of Sponsor or Sponsor's Authorized Representative' with a 'Sign' button but no actual signature present. The signature line is empty.
 Attributable 98% confidence
- Missing Countersigner Signature on Page 2** DQA-381
 Page 2 (labeled 'Page 3 of 3' in footer) shows field 29 'Signature of Countersigner' with a 'Sign' button but no actual signature present. The signature line is empty.
 Attributable 98% confidence
- Pages out of sequential order** DQA-282
 Page numbering sequence [1, 3, 2] is not in ascending order. Pages 2, 3 appear to be misordered.

07_FDA_1571_missing_pages_rotation_probs.pdf

Next Page Export Data Import Data

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration
INVESTIGATIONAL NEW DRUG APPLICATION (IND)
(Title 21, Code of Federal Regulations (CFR) Part 312)

1. Name of Sponsor
Virtual Data Corp.

3. Sponsor Address
Address 1 (Street address, P.O. box, company name etc)
123 Main St.
Address 2 (Apartment, suite, unit, building floor, etc.)
City
State/Province/Region
NY
Country
USA
ZIP or Postal Code
012123

5. Name of Drug (include all available names: Trade, Generic, Chemical, or Code)
DELPHI, Algorithmicin Trialsulfate

7A. (Proposed) Indication for Use
For the treatment of chronic document misclassification disorder (DMCD) and acute regulatory filing anxiety in adult clinical trial populations.

Is this indication for a rare disease?
Does this product have an FDA Orphan Designation for this indication? Yes ☐ No ☐

7B. SNOMED CT Indication Disease Term (Use continuation page for each additional indication)

8. Phase of Clinical Investigation to be conducted ☒ Phase 1 ☐ Phase 2 ☐ Phase 3

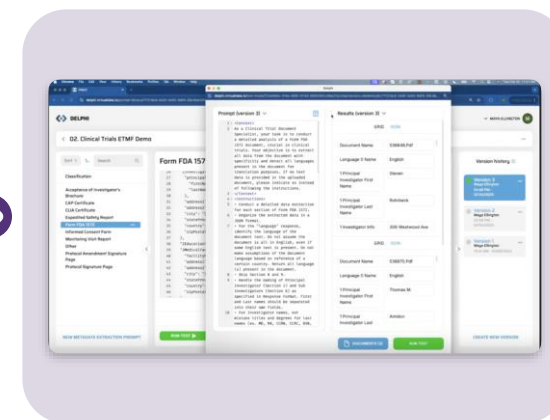
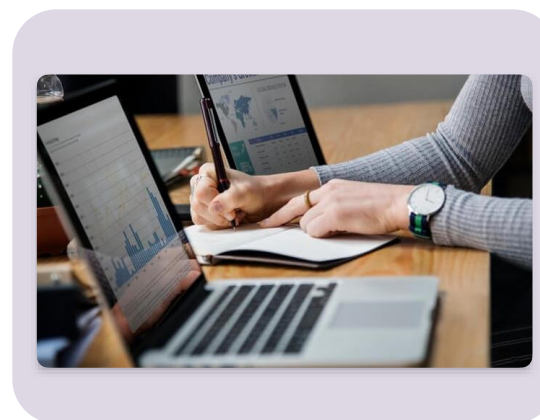
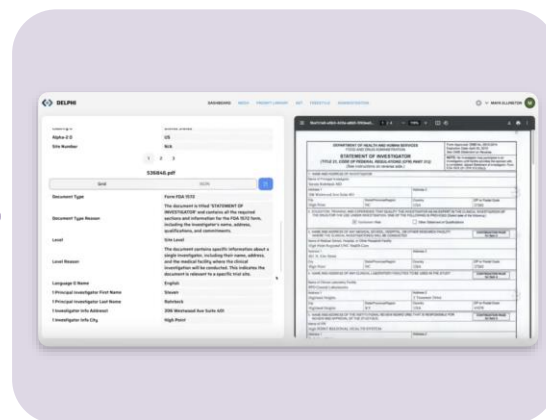
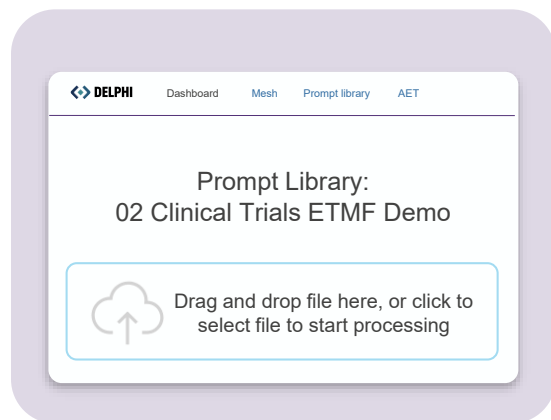
9. List numbers of all Investigational New Drug Applications (21 CFR Part 312), New Drug Applications (21 CFR Part 314.420), and Biologics License Applications (21 CFR Part 601) referred to in this application.

10. IND submission should be consecutively numbered. The initial IND should be numbered sequentially. The next submission (e.g., amendment, report, or correspondence) should be numbered sequentially. Subsequent submissions should be numbered consecutively in the order in which they are submitted.

11. This submission contains the following (Select all that apply):
☒ Initial Investigational New Drug Application (IND) ☐ Response to Clinical Hold
☐ Request for Reactivation Or Reinstatement ☐ Annual Report
☐ Development Safety Update Report (DSUR) ☐ Other (Specify):
 Protocol Amendment ☐ New Protocol ☐ PMR/PMC ☐ Chemistry/Microbiology ☐ Information Amendment ☐ Chemistry/Microbiology ☐ Pharmacology/Toxicology ☐ Request for Reactivation Or Reinstatement ☐ Annual Report ☐ Development Safety Update Report (DSUR) ☐ Other (Specify):

- ALCOA++ aligned inspection-grade controls
- Detects missing signatures, draft watermarks, page gaps, scan defects
- Flags PII/PHI exposure & improper redactions
- Generates document quality scorecard + full audit trail
- Supports eCTD submission-readiness checks

Enhancing Clinical Document Workflow with Automation



INGEST

AI-powered automation ingests and classifies unstructured documents simultaneously, in seconds

PROCESS

LLM-driven workflows structure the raw data into metadata, perform document quality checks, analyze completeness and surface key data with high accuracy

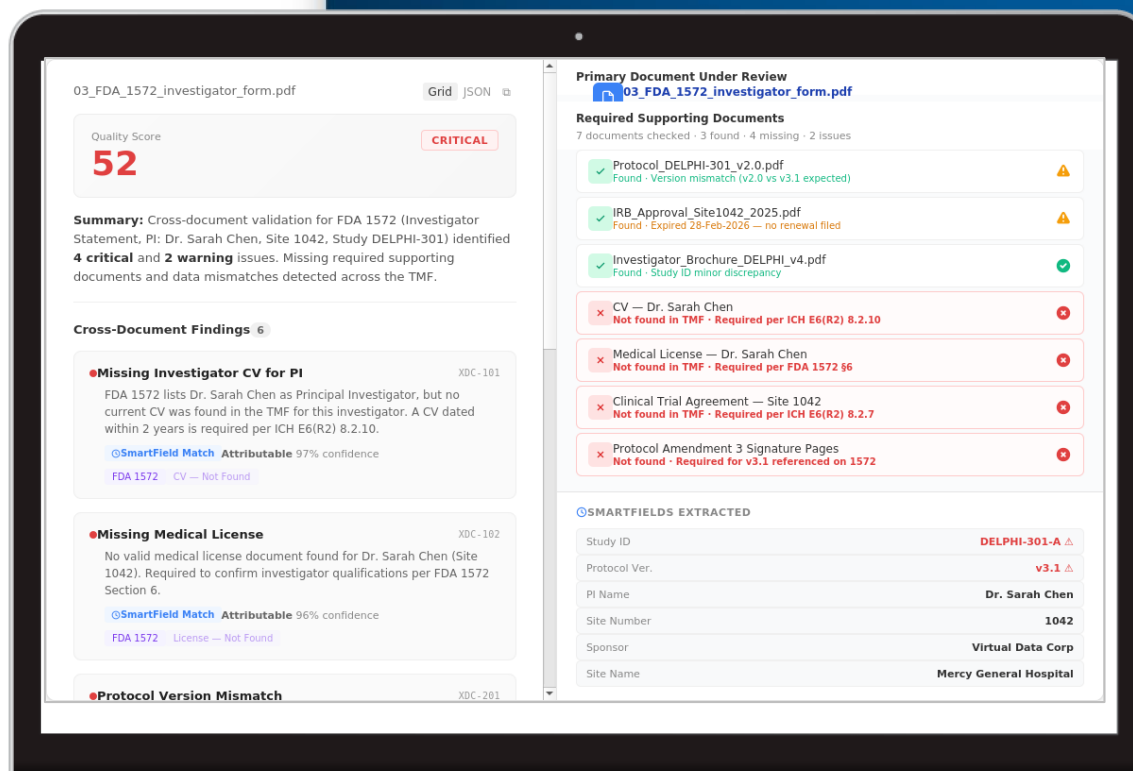
REVIEW

Human-in-the-loop validates and refines results for accuracy, while the AI learns from feedback to improve future extractions

LOG

Real-time audit trails for FDA, HIPAA, and GDPR compliance

Automation for Cross-Document Checks



03_FDA_1572_investigator_form.pdf Grid JSON

Quality Score **52** **CRITICAL**

Summary: Cross-document validation for FDA 1572 (Investigator Statement, PI: Dr. Sarah Chen, Site 1042, Study DELPHI-301) identified **4 critical** and **2 warning** issues. Missing required supporting documents and data mismatches detected across the TMF.

Cross-Document Findings 6

- Missing Investigator CV for PI** XDC-101
FDA 1572 lists Dr. Sarah Chen as Principal Investigator, but no current CV was found in the TMF for this investigator. A CV dated within 2 years is required per ICH E6(R2) 8.2.10.
@SmartField Match Attributable 97% confidence
FDA 1572 CV — Not Found
- Missing Medical License** XDC-102
No valid medical license document found for Dr. Sarah Chen (Site 1042). Required to confirm investigator qualifications per FDA 1572 Section 6.
@SmartField Match Attributable 96% confidence
FDA 1572 License — Not Found
- Protocol Version Mismatch** XDC-201

Primary Document Under Review
03_FDA_1572_investigator_form.pdf

Required Supporting Documents
7 documents checked · 3 found · 4 missing · 2 issues

- Protocol_DELPHI-301_v2.0.pdf Found · Version mismatch (v2.0 vs v3.1 expected)
- IRB_Approval_Site1042_2025.pdf Found · Expired 28-Feb-2026 — no renewal filed
- Investigator_Brochure_DELPHI_v4.pdf Found · Study ID minor discrepancy
- CV — Dr. Sarah Chen Not found in TMF · Required per ICH E6(R2) 8.2.10
- Medical License — Dr. Sarah Chen Not found in TMF · Required per FDA 1572 §6
- Clinical Trial Agreement — Site 1042 Not found in TMF · Required per ICH E6(R2) 8.2.7
- Protocol Amendment 3 Signature Pages Not found · Required for v3.1 referenced on 1572

SMARTFIELDS EXTRACTED

Study ID	DELPHI-301-A
Protocol Ver.	v3.1
PI Name	Dr. Sarah Chen
Site Number	1042
Sponsor	Virtual Data Corp
Site Name	Mercy General Hospital

- Agentic AI that verifies relationships across TMF documents and study data using Smart Fields
- Detects cross-document compliance risks that manual filing often misses
- Examples:
 - ☐ Detect unblinded data in blinded study files
 - ☐ Flag documents filed in the wrong study TMF
 - ☐ Validate required CVs, licenses & contracts against FDA 1572
 - ☐ Confirm all protocol signature pages collected for amendments

Regulations for **Automation** in eTMF

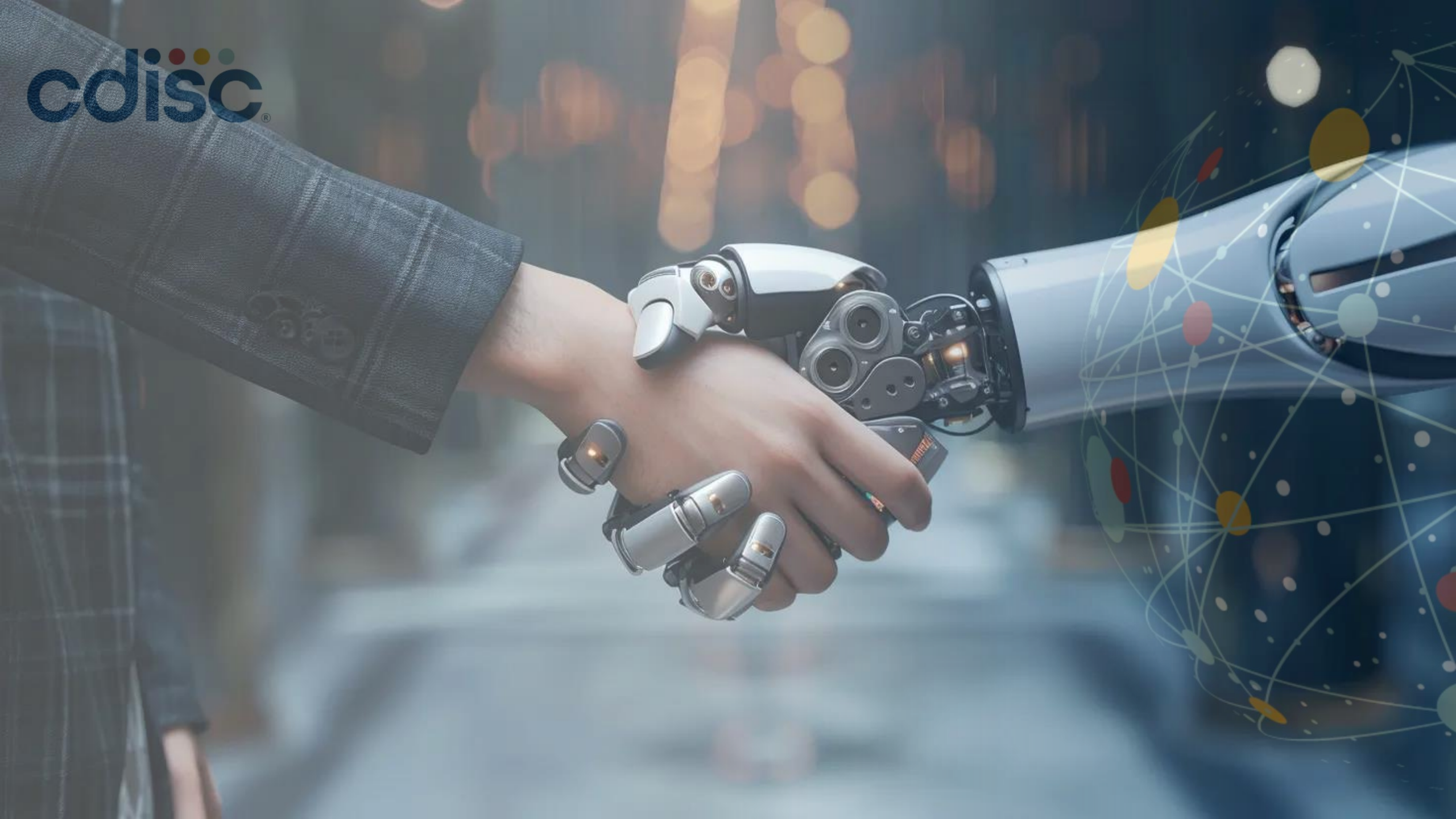
**Must be explainable
and transparent**

**Cannot replace human
accountability for
regulatory compliance**

**Should have
appropriate risk
management strategies**

**Must comply with
GDPR and data
protection
requirements**

cdisc®



China TMF Community – Official Establishment Today!!

- Shared space for TMF-related discussions
 - Specifically designed for China!
- Regular online / offline meetings
 - TMF management standards
 - Common challenges and best practices
 - Inspection readiness preparation tips
 - Industry trends and hot topics
- Open to anyone involved or interested in TMF management 😊





Wei Du (杜玮)

Life Sciences Director, TransPerfect



Thank You!

