MetaLark /AI Services Layer
Content Extraction Report Results

Customer	T-Mobile
Environment	Production · Extraction Run: 29 May 2026 / delta catch up June 11, 2026
AI Services Go-Live	11 June 2026
Prepared By	DG · Senior Technical Services Engineer

Executive Summary

The Bulk AI Extraction process was executed across the majority of Active/Inactive and Published/Unpublished Courses and Nuggets in T-Mobile's Production LMS environment to extract text content to enable AI-powered features including metadata generation, skills mapping, summary detail, assessment creation, and search enhancement.

Content that was excluded per the request of T-Mobile, or known to not produce meaningful extractions:
Items with less than 6 users assigned, HTML / Text type Nuggets, URL or Link type Nuggets, Foreign Source objects (except Xyleme), Shared Replica Courses / Nuggets, and Courses / Nuggets missing assets.

The extraction pipeline was evaluated across two phases: (1) outright extraction success or failure, and (2) quality review of successfully extracted content. The results demonstrate strong performance with a 97.59% success rate, and informed targeted improvements to the extraction process.

Content that failed outright points to some error in the actual extraction process itself. These indicate issues that will first be reviewed by OnPoint and addressed where possible, though sometimes these failures are due to incorrectly built or unsupported content items as well.

7,003	6,834	169	97.59%
Learning Objects Evaluated	Successfully Extracted	Failed Extracts	Extraction Success Rate

Extraction Success

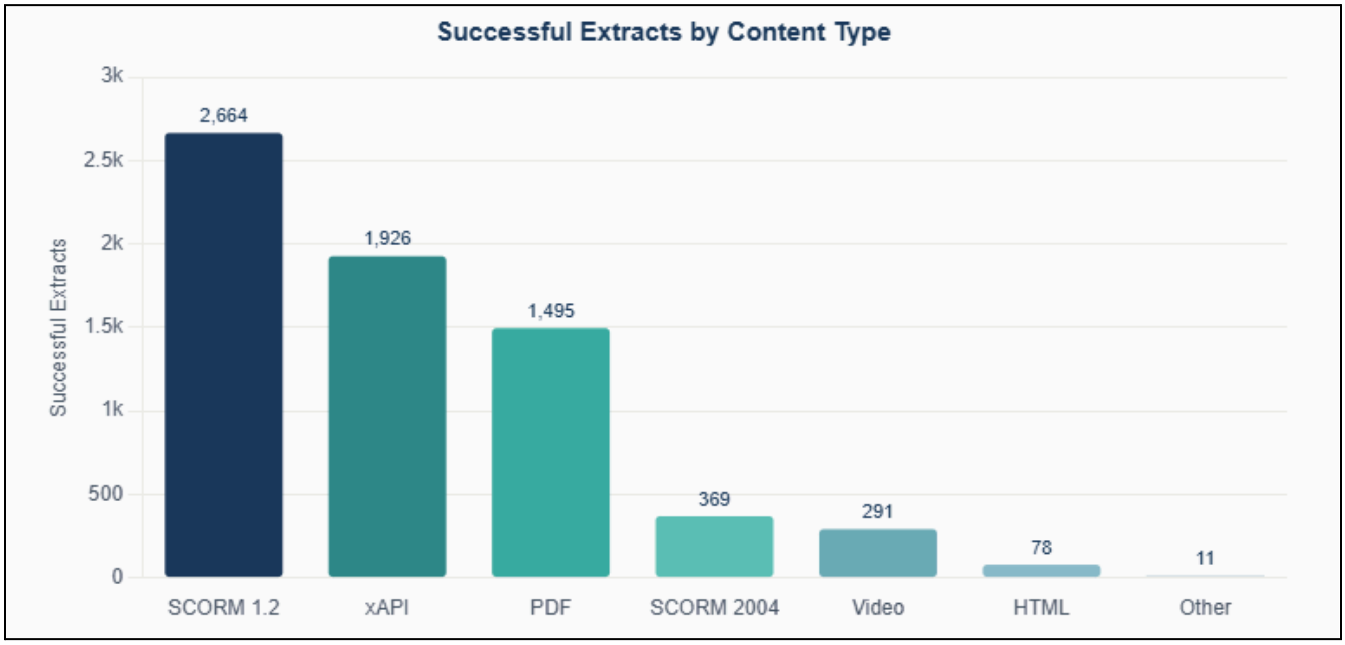
Overall, 7,003 items were processed with 6,834 successfully extracted. Courses accounted for the majority of items at 5,101, achieving a 97.43% success rate, while Nuggets had a slightly higher success rate of 98.00% across 1,902 items.

Object Type	Total	Successful	Failed	Success %
Active/Inactive Courses	5,101	4,970	131	97.43%
Active/Inactive Nuggets	1,902	1,864	38	98.00%

Object Type	Total	Successful	Failed	Success %
Total	7,003	6,834	169	97.59%

Content Type Breakdown

Successful extracts span a wide range of course and nugget formats. SCORM 1.2 and xAPI content represent the largest volume, followed by PDF nuggets. Articulate *Rise* and Articulate *Storyline* are the dominant authoring tools across the extracted course library.



Course Types — Successful Extracts

Packaged Content	Count
SCORM 1.2	2,664
xAPI	1,926
SCORM 2004	369
Other (non-tracking HTML)	11

Course Authoring Tools — Successful Extracts

Authoring Tool	Count
Articulate Rise	3,109
Articulate Storyline	1,642
Generic	137
Adobe Captivate	38
ELB Learning Lectora	18
iSpring	12
gomo	11
dominKnow	3

Nugget Types — Successful Extracts

Media Type	Count
PDF	1,495
Video	291
HTML (PDF type nugget with HTML asset linking to OPDoc)	78

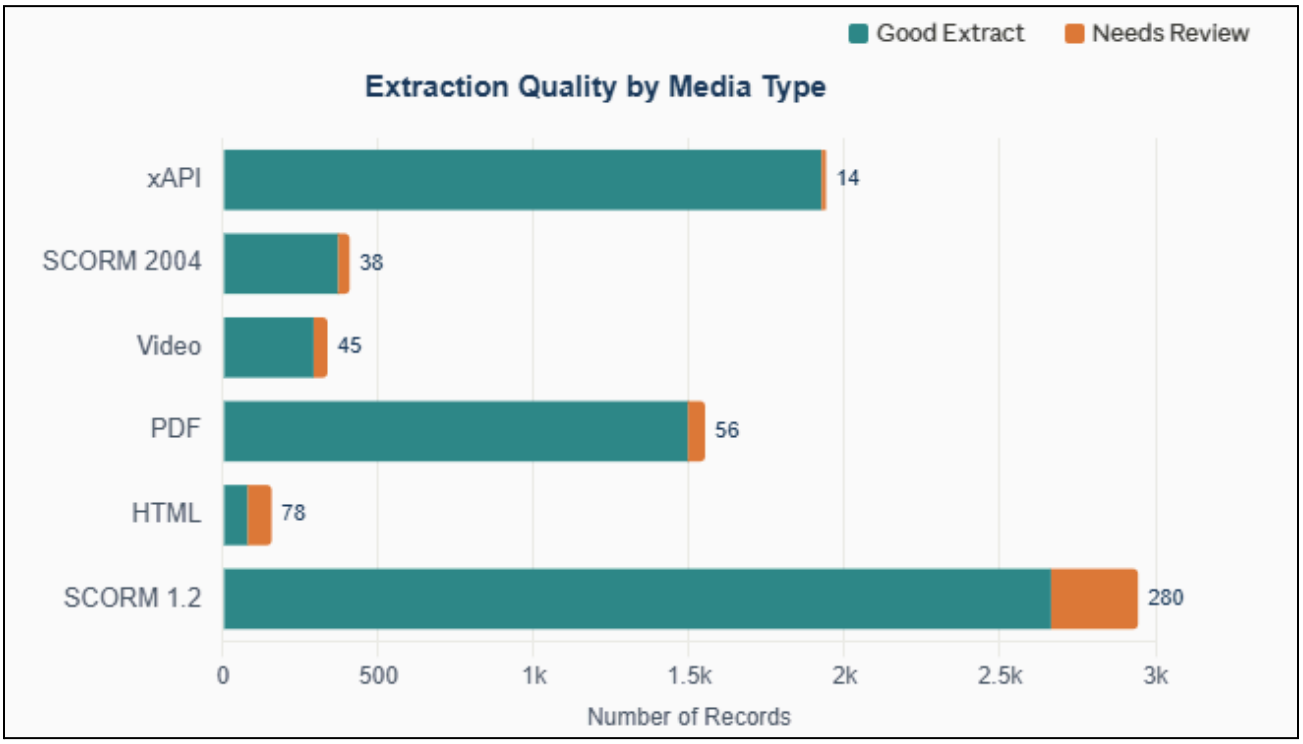
Extraction Quality Factors

Of the 6,834 total extracts created, 511 (7.48%) were flagged as Needs Review based on AI-driven quality analysis. Flagged items are assessed against defined criteria — this does not necessarily indicate a content problem, but identifies items warranting closer inspection.

6,834	6,323	511	7.48%
Total Extracts Created	Clean Extracts	Needs Review	Review Rate

Needs Review — By Content Type

As you can see in the chart below, the majority of the items that need review fall under the SCORM 1.2, SCORM 2004, and xAPI Content Types. Non Tracking HTML based content provides the next largest set of items needing review. This includes those html based nuggets that load external data or videos into the Nugget frame. This also includes HTML nuggets that reference OPDoc documents that are under current review. PDFs that were determined to “need review” were largely image only pdfs. The upcoming update to the extraction process will accommodate for these images only items and can successfully extract their content on a reprocess.



Needs Review Analysis

The following describe the reasoning behind items flagged as ‘Needs Review’. Minimal Content is by far the dominant cause, accounting for over 88.65% of flagged records, and largely reflects content that simply contains limited extractable text rather than a technical failure. Extracts that successfully ran but produced a blank file are identified by a Log Reason of "Empty Content Array." This typically indicates content that contains no spoken audio or loads all content from an external source.

Root Cause	Count	% of Needs Review
Minim al Content	453	88.65%
Empty Extract	58	11.35%

Minimal Content / Empty Extract — Pattern Breakdown

Of the 511 Minimal Content and Empty Extract items, 295 were reviewed in depth, with the following findings:

- **155 items** consisted of externally linked content (dispatch) or externally linked videos, which yielded minimal extracted content. As these sources are inaccessible during extraction, only text available in the course package was captured.
- **78 items** were PDF-type Nuggets containing HTML assets that reference OPDoc documents. As previously noted, this is not yet supported and the team is currently investigating approaches to extracting the underlying PDFs referenced within OPDoc.

- **41 items** were videos containing sound effects only, with no spoken audio, resulting in an empty extract.
- **21 items** were PDF-type Nuggets containing infographics with minimal text, these will be re-run once the upcoming extraction process release is deployed which contains the enhanced image handling that is expected to yield additional extracted data.

Key patterns identified:

Review Reason	Count
Courses with externally linked video	100
PDF type nugget w/ link to OPDoc	78
Externally linked content (dispatch)	55
Video with sound effects only (no words)	41
PDF infographics with minimal text	21

Remediation & Recommended Actions

Immediate Actions

Issue	Recommended Action	Owner	Status
Image-based PDFs	Extraction process updated to extract text from Image-based PDFs.	OnPoint	Upcoming Release
Special characters in filenames/names	Extraction process updated to handle special characters. Corrected in this run.	OnPoint	Complete
Objects with missing assets	Content with no attached asset should be reviewed and corrected, otherwise it will be excluded from the bulk extract.	T-Mobile	Complete

Strategic Improvements

Improvement	Status
Extraction of Nuggets that reference links to OPDoc documents	Under Review
Image-based PDF OCR extraction	Upcoming Release
Special character support (filenames & object names)	Complete
Skip unsupported media files (m3u8)	Complete
External content crawling / browser extraction	Under Consideration
Standardize authoring on Articulate Rise, Articulate Storyline and PDF	T-Mobile action

Next Steps

Action	Description	Owner	Status
Review AI-generated metadata	For all "Good" extracts, review and validate AI-generated metadata in the platform.	T-Mobile	Pending Review
Continue evaluating extracts flagged for needs review	OnPoint reviewed a large portion of these, but there are others that can still be reviewed to justify why the extracts were flagged for review. A separate Excel document will be provided with a tab for “Needs Review”. Items that have not been reviewed will have an empty “reviewReason” column.	T-Mobile	Pending Review
Rerun for image based extraction	Once the extraction process has been updated to the upcoming release, all “Needs Review” and “Failed Extracts” content will be reprocessed, as the enhanced image handling is expected to yield additional extracted data for any courses or nuggets containing images. This may result in content currently flagged as Minimal Content or Empty Extract to be resolved.	OnPoint	Upcoming Release / In Review

OPDoc Document Extraction	OPDoc Documents can be extracted to provide title, description, and metatag suggestions. While TMO previously chose not to move forward with this functionality, it remains a viable option open for reconsideration at any time.	T-Mobile	Pending Review
Apply AI-generated Metatags	AI-generated metatags will be applied only when the suggested tag does not already exist as a manually assigned metatag, as agreed upon by the TMO team.	OnPoint	Complete
Update Content Descriptions based on AI-generated summary	All successfully extracted content will be updated with the AI-generated Summary's Overview and Main Topics sections specifically, as agreed upon by the TMO team.	OnPoint	Complete

Final Status

✓ AI Enablement Status: COMPLETE