

2026 Excellence in Technology Awards Program

ENTRY CATEGORY OR CATEGORIES:

Future of Work:

X	Best Advance in AI for Business Impact
---	--

Entry Overview:

Entry Title	MetaLark.ai: A Content Intelligence Toolkit for L&D Teams
Entering Company	OnPoint Digital, Inc.
Date	September 4, 2026
Category	Best Advance in AI for Business Impact
Brief Description of Entry (2-3 Sentences)	MetaLark.ai is a Content Intelligence Toolkit that makes an existing learning library legible, searchable and measurable. It can ingest digital content resources from any LMS, LXP or cloud repository — including SCORM, xAPI, cmi5 and AICC packages whose instructional text is unreadable by conventional search tools and language models — and then generates a full-text transcripts & JSON+MD artifacts used to generate AI-validated metadata, content summaries, inferred skills and quality diagnostics for every asset. MetaLark evaluated 7,003 learning objects and successfully extracted 6,834 of them at a 97.59% success rate in 36 hours, against an estimated 1,750 staff hours for equivalent manual review.



Company Background

Headquarters	Savannah, Georgia USA
Year Founded	2002
Revenue	\$5M+
Employees	25
Global Scale	Global customers with operations in the USA & New Zealand

Customers / Output	Enterprise learning and HCM customers across telecom, financial services, healthcare, manufacturing, government and professional associations. Services offered: learning platform delivery (OnPoint LMS/LXP & CellCast Mobile Learning), content engineering and migration services, and AI content intelligence services delivered through MetaLark.ai. Example referenced customer supporting this entry: a Fortune 100 wireless carrier based in the USA.
Industry	Enterprise Software for Learning / HCM
Website	http://metalark.ai

Mission. OnPoint Digital builds technology that lets learning organizations act on evidence about their own content. MetaLark.ai extends that mission to the largest and least examined asset most L&D functions own — the content library already in place — on the premise that an organization cannot govern, modernize or make AI-ready a library it cannot read.

Value Proposition (based on judging criteria 1: Value Proposition)

The business need: L&D teams have libraries they cannot see inside of

Most learning organizations direct nearly all of their content budget toward producing new material, while their largest untapped assets sit unexamined — the library of learning content already resident in their LMS, LXP and all corporate cloud repositories. And much of that multi-domain library is typically illegible to the organization that owns it as content is often locked inside packaged courseware bundles. Catalogs and reporting tools can list content titles, general descriptions and available metatags (if provided by an Admin) as well as details on the content owners and intended audiences, but they cannot fully describe what is actually inside every managed learning object thus any inventory is incomplete.

The reason is structural rather than administrative. Standards-based courseware is not machine-readable. Instructional text in an AICC, SCORM 1.2, SCORM 2004, xAPI or cmi5 package sits inside a zipped archive of fragmented HTML, XML manifests and JavaScript that no enterprise search tool, LMS report or commercial Large Language Model can address. Video content exists only as spoken audio, not the imagery displayed in motion frames. A large share of enterprise PDFs exist as image-only scans with no extractable text layer. Each authoring tool packages its output differently, so a

library assembled over two decades from several tools is effectively a set of sealed containers.

The consequence is that a set of ordinary business questions has no answer. Which assets mention a specific product name, brand term or regulated word or phrase? What already exists on a topic before a new build is commissioned? Which content maps to which skills? What is duplicated, and what is stale? Absent a way to read inside the assets, each of these becomes a manual review project priced in person-months, and the answer expires the moment the review ends.

MetaLark.ai addresses all of these challenges with a proven, automated process and highly refined and validated tooling. It is positioned as a diagnostic instrument for the content ecosystem — *a Digital MRI for your Learning Content* — rather than as another authoring or delivery tool. The premise is that high-dollar decisions about learning strategy should not be made without first scanning what is inside the library, in the same way that surgery is not scheduled without imaging the patient.

How the solution works

MetaLark operates in three stages, and the operating simplicity is deliberate: the process is designed to be run by an LMS administrator or content manager rather than by a data team.

- **Ingestion.** MetaLark connects to the source systems of record — LMS or LXP via API, SharePoint, Google Drive, Amazon S3 — or accepts files dragged and dropped directly into the application. No content is moved out of governance nor replaced.
- **Scan and diagnose.** A machine-learning parsing engine deconstructs every asset layer by layer: unpacking course archives, transcribing media, and extracting text from documents and images. For each object it produces a full-text transcript, an AI-authored title and description, a structured summary with an overview and main topics, inferred skills mapped to a standard ontology such as ESCO or to the customer's own framework, AI-generated assessment items and question banks, and diagnostic quality flags. Scanning runs at under a minute per item.
- **Prescribe and deploy.** Every object is assessed against a four-part disposition framework — Retain, Refresh, Retire, Reuse — and the results can be exported to a new LMS or LXP, an enterprise language model, or an authoring tool in an AI-ready, skills-aligned form.

The disposition framework is what converts an inventory into a decision set. “Retain” identifies high-value, accurate, evergreen content confirmed in excellent condition and safe to deploy with confidence. “Refresh” identifies relevant content showing signs of

age, and indicates which items need updating and where. “Retire” identifies outdated or redundant content that should be removed to reduce clutter, risk and compliance exposure. And “Reuse” identifies strong assets with untapped potential that can be reimagined for new audiences, formats and contexts.

How the solution is used today

MetaLark is used by the distinct roles that touch learning content — LMS administrators, content publishers, learning consultants and systems integrators who are often tasked with performing content inventories and migration efforts for larger enterprise accounts. Four use patterns account for most current deployments.

1. Full-text search with pinpoint location

A team needs to deploy updated training quickly. Existing content may well cover the right topics, but confirming that means manually opening repositories to check relevance, freshness and authoring-tool compatibility, and any one of those checks can turn out to be a dead end. MetaLark searches the entire unified library in milliseconds, filtered by content type, authoring tool, publish date or metadata, and surfaces all exact and semantic matches — including many assets no one remembered even existed. Typical review sessions happen in minutes rather than hours or days, and previously invisible content re-enters circulation or gets dealt with appropriately.

2. Consolidation of scattered repositories into one canonical library

Following an acquisition or a reorganization, a learning function faces half a dozen content repositories — multiple legacy LMS platforms, Microsoft SharePoint, Google Drive — with no consolidated picture of what the combined organization now owns. Cross-referencing disparate systems by hand to build a unified inventory is impractical at enterprise volume. MetaLark ingests all of it and produces a single searchable library complete with auto-generated metadata, assessments and skills mapping. A fully consolidated library is typically ready within a day or two, which allows strategic content planning and executive-ready reporting to begin immediately rather than after a multi-month audit.

3. Exposing the skills already covered by existing content

When leadership commits to a skills-based talent model, the L&D function is asked which content supports which skills. The usual position is that metadata is incomplete, no skills engine has been funded, and no one can answer. MetaLark auto-generates skills for every content item during extraction, mapping findings to a standard ontology such as ESCO, to the organization's custom skills framework, or to both, with no manual

tagging required. Skills are therefore mapped and available before any search begins, and before a separate skills platform investment is made.

4. Platform migration and consolidation

A team migrating from legacy platforms to a consolidated system must decide what should make the move. Where the last content audit is years old, outdated content is hard to identify, metadata is patchy, and coordination across multiple source platforms is disorderly. MetaLark consolidates all repositories into a single review environment without requiring administrators to log into the source systems, flags each item to retain, refresh, retire or reuse, organizes content into administrator-defined collections, and exports through a secure API or pre-defined, vendor-specific import templates for a clean, structured migration effort that often saves days of effort. A chaotic and expensive scramble becomes a planned, budgeted project with a defensible scope.

Customer example: Fortune 100 Wireless Carrier

A well-known Fortune 100 wireless carrier's learning content ecosystem had grown by accretion over two decades, with content arriving from different business units, agencies, acquired smaller and market-specific carriers and authoring standards. By 2026 the production environment held 7,003 in-scope learning objects — 5,101 courses and 1,902 content nuggets — authored across eight different tools. The catalog listed titles and audiences but could not describe what was inside any of them, and applied metadata was widely regarded as incomplete and inaccurate.

Three concurrent demands made that untenable. An enterprise skills agenda had been committed to, and the selected skills intelligence vendor required content metadata that was present and accurate in order to map earned skills to learning; metadata for more than a third of in-scope content was absent or wrong, which delayed the skills strategy on what was fundamentally a data problem. Separately, legal mandates required the wireless carrier to identify, isolate and document every instance of specific language deemed inappropriate in a charged regulatory environment — an obligation with no manual answer across 7,003 files. Finally, every new content request was being judged on its own merits, because nobody could say what already existed covering the same ground, which meant the function was funding duplication it had no way to see.

MetaLark was connected directly to the carrier's production LMS environment, unpacked every course package, transcribed all media, and extracted text from each layer of every digital resource. The bulk run executed on May 2026 with a delta catch-up on mid-June 2026, when MetaLark's AI content services were formalized in the environment. It evaluated 7,003 objects and successfully extracted 6,834 at a 97.59% success rate. Full results appear in Section 5.

The material outcome was not better courseware. It was an operating model in which the L&D function can answer questions about enterprise content with evidence in minutes rather than weeks — pinpointing brand-specific or regulated language inside any file, supplying the enterprise skills platform with trusted metadata, and exposing duplication that can reshape content investment decisions. All new content uploaded to the wireless carrier's LMS and third-party content management platform is now scanned and indexed automatically, so tagging, skills alignment and trend data stay current rather than degrading between audits.

Product or Program Innovation (based on judging criteria 2: Innovation)

The breakthrough: making packaged courseware machine-readable

The core innovation is narrow and consequential. MetaLark reads the instructional content inside standards-based courseware packages, which no commercial search tool and no frontier language model can do unaided. For the wireless carrier, 4,959 of the 7,003 in-scope objects were packaged courses — 2,664 in SCORM 1.2, 1,926 in xAPI and 369 in SCORM 2004 — in which the text sits inside zipped archives of fragmented HTML, XML manifests and JavaScript addressable by no search tool. A further 291 objects were videos whose content existed only as spoken audio, and 1,495 were PDFs including many image-only files with no text layer at all.

Solving that parsing problem changes the unit of work. Content review stops being the act of opening and reading individual assets and becomes the act of querying a structured corpus. That is a difference in kind rather than degree, and it is what makes every downstream capability possible — search, skills inference, similarity detection, compliance research, and any form of agentic retrieval over learning content.

It also makes the alternative visible for what it is. Manual review of an enterprise library at a conservative fifteen minutes per object — enough to launch it, skim it and record findings — amounts to roughly 1,750 person-hours across 7,003 objects, or about eleven person-months of instructional designer and subject-matter-expert time diverted from their actual work. It leaves no durable artifact, so the library begins drifting again the day the review ends. And sampling cannot deliver the consistency across every file that a general counsel requires.

Simplicity and time to value

The solution is engineered so that value arrives before a project is chartered. Prospective users can try a free scan at no cost, uploading up to twenty items of any

size and complexity into a private AWS instance for a thirty-day evaluation, which produces real diagnostic output on the organization's own content rather than on a demonstration data set. At production scale, ingestion requires an API connection or a folder rather than a data engineering effort; scanning runs at under a minute per item; and a fully consolidated, searchable library across multiple repositories is typically ready in a day or two. The wireless carrier's end-to-end review of 7,003 objects completed in 36 hours of processing.

Advanced functionality

The diagnostic capabilities extend well past inventory and tagging.

- **AI-generated metadata at scale.** Rich metatags, titles and descriptions generated in seconds per object, with structured summaries carrying an overview and key takeaways.
- **Full-text content encyclopedia.** Every asset searchable to its text, across all content types and source systems simultaneously.
- **Pinpoint location and deep linking.** Matches are addressable to the slide, transcript segment, video frame, displayed graphic or embedded document, with deep links that jump directly to the relevant term inside the content.
- **Embedded content player.** Content launches for immediate review inside MetaLark, positioned directly at the search result, without returning to the LMS.
- **Skills inference and mapping.** Skills are inferred per object and mapped to the ESCO ontology, to customer-specific taxonomies, or to both.
- **AI-generated assessments.** Question banks and assessment items generated from the extracted content.
- **Analytical reporting.** Content analysis by type and authoring tool, extraction and failure diagnostics, Bloom's taxonomy analysis at both library and item level, and advanced insights covering content similarities, thematic clusters and overlapping titles.
- **Cross-domain organization.** Content from multiple sources organized into administrator-defined collections with status and action markers reflecting the “Retain, Refresh, Retire and Reuse” dispositions.

Flexibility, customization and integration

MetaLark is designed to coexist with the systems an organization already runs rather than to displace them. It ingests SCORM 1.2 and 2004, xAPI and cmi5 packages; documents and media including PDF, HTML, Microsoft Word, Microsoft PowerPoint, image files, MP4 and MP3 rich media, and VTT transcript files; and output from eight authoring tools including Articulate Rise and Storyline, dominKnow One, iSpring Suite,

Adobe Captivate, ELB Lectora, Gomo Learning, Adapt Builder and custom packagers. Integration runs through LMS and LXP APIs, SharePoint, Google Drive, Dropbox and Box, and Amazon S3 buckets, with enterprise SSO, ESCO skills ontology mapping, custom tag and skills taxonomies, AWS-hosted cloud infrastructure and SOC 2 certification. Agentic AI access is supported via a universal API key as well as via a MetaLark MCP (Model Context Protocol) Server, and extracted content is chunked into a vector database so that downstream systems can retrieve at the keyword or skill level rather than launching whole objects.

The wireless carrier's deployment demonstrates the coexistence. MetaLark supplies the content truth layer while a separately selected skills intelligence vendor supplies the approved ontology, talent model and growth recommendations. Two AI systems from different vendors operate as a consolidated pipeline, which is only possible because MetaLark is scoped as a content intelligence layer rather than as a platform seeking to own the talent architecture.

Pricing model & ROI calculator

Pricing for MetaLark is based on a transactional charge that's assessed to each content item being processed – no matter what the actual size of that content turns out to be. Most web-friendly digital files range in size from 2-3 MB on the low end of the scale but might be 250MB or 300MB at the upper end of that scale*. Teams simply purchase a bank of processing credits and the system increments down as they are used. Virtually all features and extracts are covered for the price of each processing credit and discounts on a sliding scale are realized based on the total number of credits purchased. A free scan covering up to twenty items on a thirty-day evaluation in a private instance carries no cost. A quick-quote calculator and an ROI calculator are both publicly available, so a prospective buyer can size a project and build an internal business case before speaking to a salesperson. That matters in this category specifically, because the buyer is usually funding a content audit out of a budget that has no line item for one, and needs a defensible number against the person-hour alternative.

* = Large rich media content items exceeding one hour of runtime require an additional processing credit.

Development methods and client service approach

Four practices govern how the technology is deployed, each of them evidenced in delivered client outcomes rather than asserted.

- **The AI defers to human judgment.** AI-generated metatags are applied only where no manually assigned tag already exists unless directed by a customer. Human-authored

metadata is preserved by design. No object is automatically retired, altered or republished; disposition recommendations arrive as flags on a review queue with named human owners.

- **Quality is measured and disclosed rather than assumed.** Extraction is evaluated in two conditions — outright success or failure, then quality review of what succeeded. Flagged items are root-caused and reported to the client with counts and patterns, and a remediation plan with named owners and status is issued. For the wireless carrier, 511 objects were flagged for review and 295 were examined individually with the patterns documented.
- **Data stays governed.** Processing runs in the client's dedicated environment under SOC 2 controls with enterprise SSO. Client content is not used to train external AI models, under strict data residency and retention terms.
- **Findings drive product development.** Diagnostic gaps identified in customer extraction programs become scheduled engineering work with published status. Image-only PDF OCR extraction, special-character handling in filenames and object names, unsupported media skipping and external content crawling each entered the roadmap directly from client extraction analysis, and flagged and failed objects are reprocessed as the pipeline improves.

Market positioning

MetaLark occupies a position that the learning technology market has largely left vacant. Authoring tools create content. Learning platforms deliver it. Skills platforms model the talent side. None of them read what is already in the library. As organizations move toward enterprise language models and agentic search over corporate knowledge, that gap becomes the constraint: an AI-ready learning ecosystem cannot be built on a library nobody can read, and feeding a retrieval system unexamined content propagates outdated and non-compliant material rather than retiring it. MetaLark is positioned as the step that has to happen first, and it is deliberately scoped to that step.

Unique Differentiators (based on judging criteria 3: Unique Differentiators)

MetaLark is frequently compared to three adjacent classes of technology: content management and DAM systems, LMS or LXP search, and skills intelligence platforms. The differentiators below are stated against those comparisons specifically.

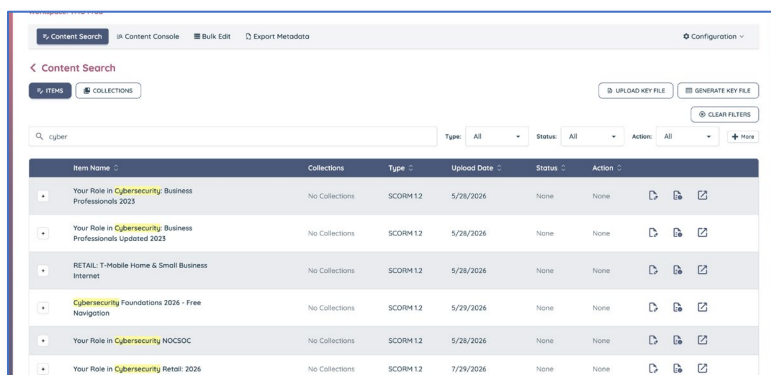
1. It reads inside the containers other tools treat as opaque

This is the foundational differentiator and the one that gates all the others. Content management systems index the metadata attached to a file; LMS search indexes titles,

descriptions and applied tags. Neither reads the instructional text sealed inside an AICC, SCORM, xAPI or cmi5 archive, inside a video's spoken audio, or inside an image-only PDF. MetaLark's parsing engine deconstructs each of those layers across eight authoring tools and four packaging standards. For the wireless carrier, that meant 4,970 of 5,101 packaged courses became searchable to their full text, which no incumbent system in that environment could do.

2. It addresses content at the concept level, not the object level

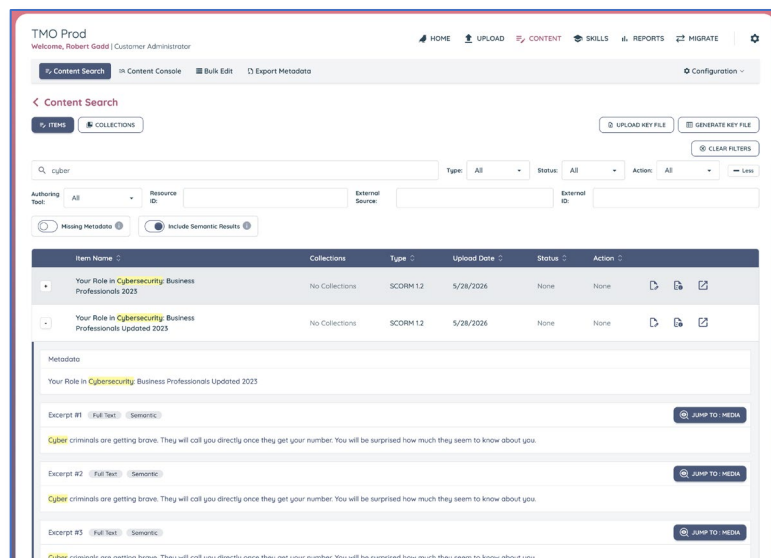
Conventional search returns a list of courses to open. MetaLark returns the location of the answer — the slide, the transcript segment, the video frame, the displayed graphic, the embedded document — with a deep link that jumps directly to the matched term and an embedded player that launches the asset positioned at that point. For compliance research this is the difference between a shortlist and a finding. For learner-facing delivery it is the basis for retrieval that serves a concept rather than a whole course.



The screenshot shows the 'Content Search' interface with a search bar containing 'cyber'. Below the search bar is a table of results. The table has columns: Item Name, Collections, Type, Upload Date, Status, and Action. The results are as follows:

Item Name	Collections	Type	Upload Date	Status	Action
Your Role in Cybersecurity Business Professionals 2023	No Collections	SCORM 12	5/28/2026	None	[Icons]
Your Role in Cybersecurity Business Professionals Updated 2023	No Collections	SCORM 12	5/28/2026	None	[Icons]
RETAIL: T-Mobile Home & Small Business Internet	No Collections	SCORM 12	5/28/2026	None	[Icons]
Cybersecurity Foundations 2026 - Free Navigation	No Collections	SCORM 12	5/28/2026	None	[Icons]
Your Role in Cybersecurity NOCSOC	No Collections	SCORM 12	5/28/2026	None	[Icons]
Your Role in Cybersecurity Retail: 2026	No Collections	SCORM 12	7/29/2026	None	[Icons]

Figure 1 – Basic Search with Highlighted Results for Matched Content



The screenshot shows the 'Content Search' interface with a search bar containing 'cyber'. Below the search bar is a table of results. The table has columns: Item Name, Collections, Type, Upload Date, Status, and Action. The results are as follows:

Item Name	Collections	Type	Upload Date	Status	Action
Your Role in Cybersecurity Business Professionals 2023	No Collections	SCORM 12	5/28/2026	None	[Icons]
Your Role in Cybersecurity Business Professionals Updated 2023	No Collections	SCORM 12	5/28/2026	None	[Icons]

Below the table, there is a section for 'Metadata' and 'Excerpt' results. The excerpt shows the following text:

Excerpt #1 [Full Text] [Semantic] [JUMP TO MEDIA](#)

Cyber criminals are getting brave. They will call you directly once they get your number. You will be surprised how much they seem to know about you.

Excerpt #2 [Full Text] [Semantic] [JUMP TO MEDIA](#)

Cyber criminals are getting brave. They will call you directly once they get your number. You will be surprised how much they seem to know about you.

Excerpt #3 [Full Text] [Semantic] [JUMP TO MEDIA](#)

Cyber criminals are getting brave. They will call you directly once they get your number. You will be surprised how much they seem to know about you.

Figure 2 – Advanced Search with Deep Links to Jump to Relevant Term in Content

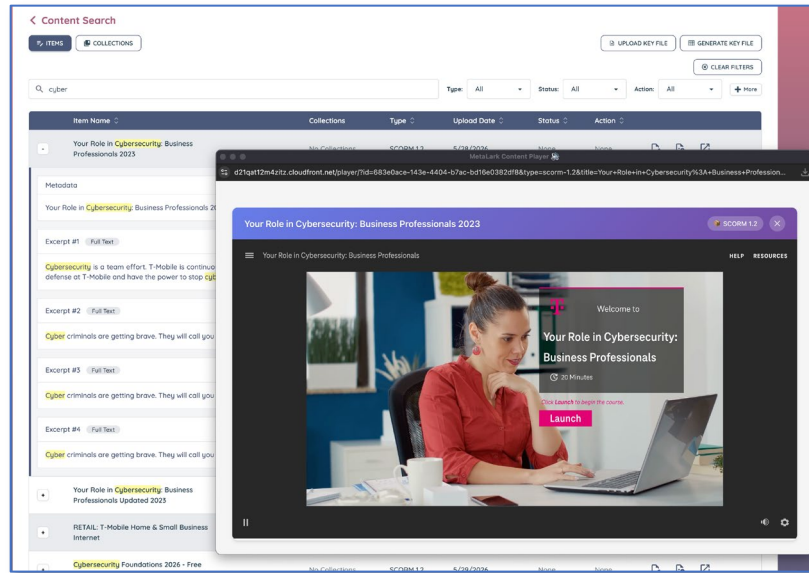


Figure 3 – Embedded Player Launches Content Directly to Search Criteria Result

3. It produces an analysis, not an inventory

An inventory tells an organization what it has. MetaLark tells it what to do. Every object is assessed against the “Retain, Refresh, Retire and Reuse” framework, and each disposition carries a specification: which items are confirmed healthy and safe to deploy, which are relevant but ageing and where specifically they need updating, which are outdated or redundant and should be removed to reduce risk and compliance exposure, and which are strong assets with untapped potential worth reimagining for new audiences and formats. The output is a triaged worklist with owners rather than a spreadsheet of file names.

4. Skills are inferred from content at extraction time, not tagged afterward

Skills platforms model skills against roles and people, then depend on content metadata to connect learning to skills. That dependency is where skills programs commonly stall, because the content metadata does not exist. MetaLark generates inferred skills from the actual text of every object during extraction and maps them to ESCO, to a customer's own framework, or to both, with no manual tagging. For the wireless carrier this produced 1,901 unique skills mapped across the library. The result is that skills are mapped and ready before any search begins and before a skills engine is deployed.

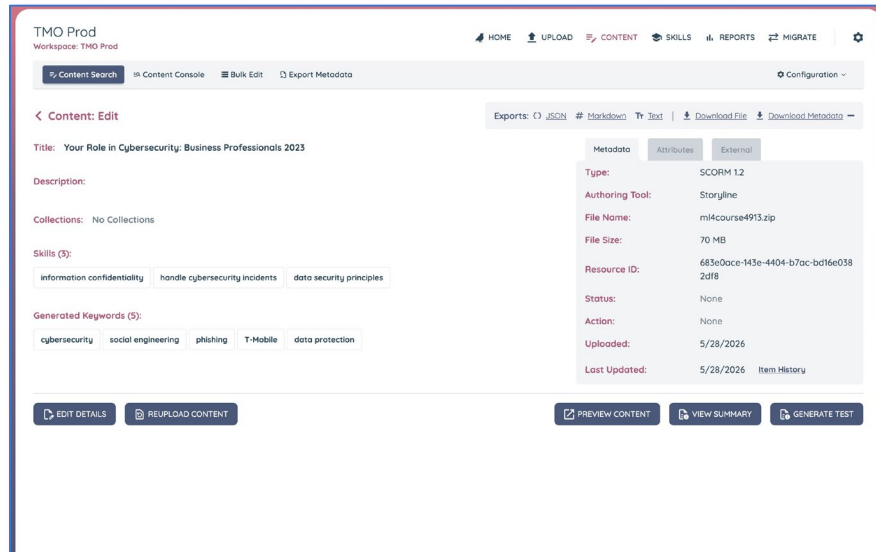


Figure 4 – Content Detail for An Extracted Resource Detailing Skills & Metatags

Skill Name	Description	Origin	Item Count
apply digital security measures	Follow simple ways to protect digital devices and content.	ESCO	17
apply information security policies	Implement policies, methods and regulations for data and information security in order to respect confidentiality, integrity and availability principles.	ESCO	8
calculate quotations for security devices services	Calculate and present quotations for the installation or maintenance services to be provided to clients, according to their door devices, costs and schedules.	ESCO	3
create a prevention security plan for the store	Design a security plan for the establishment, according to its specifications and security needs in order to prevent any shoplifting or other security threats.	ESCO	16
cyber security	The methods and best practices that protect ICT systems, networks, computers, devices, services, processes and people against unauthorised access, modification and/or denial of service of assets.	ESCO	9
data security principles	The principles of information and data security. They include, among others, confidentiality, integrity, and availability.	ESCO	14
define security policies	Design and execute a written set of rules and policies that have the aim of securing an organisation concerning constraints on behaviour between stakeholders, protective mechanical constraints and data...	ESCO	17
document security incidents in the store	Prepare documentation and specific reports of the security threats, observations and incidents, such as shoplifting, that occur in the store, to be used as evidence against the offender, if required.	ESCO	22
ensure information security	Ensure that the information gathered during surveillance or investigations remains in the hands of those authorised to receive and use it, and does not fall into enemy or otherwise non-authorized individuals' hands.	ESCO	1
ensure private property security	Make sure that gates and doors are locked, windows are closed and alarm systems are active, in order to prevent burglaries or theft.	ESCO	1

Figure 5 – Summary of Discovered & Mapped Skills for All Content

5. It composes with skills platforms rather than competing with them

MetaLark is scoped as a content truth layer, which allows it to sit upstream of whatever talent architecture an organization has chosen. For the wireless carrier, MetaLark and an independently selected skills intelligence vendor operate as a single pipeline: MetaLark supplies the vetted content and inferred skills, the skills platform supplies the approved ontology, talent model and growth recommendations. Platforms that bundle content intelligence into a broader talent suite cannot occupy that position without displacing an existing vendor decision.

6. Human-authored metadata is retained unless requested by customer

Most AI enrichment tools overwrite. MetaLark applies AI-generated metatags only where no manually assigned tag already exists, so curatorial work performed by instructional designers over years survives the enrichment. No object is automatically retired, altered or republished. This is a design commitment rather than a configuration option, and it is a material differentiator for organizations whose content owners are the reason an audit has never been approved.

7. It discloses its own limits

MetaLark reports extraction quality as a measured result rather than presenting output as uniformly reliable. Every run separates outright extraction failure from quality concerns in successful extracts, root-causes the flagged items, and issues a remediation plan with named owners and status. For the wireless carrier, 511 objects (7.48%) were flagged for review, root-caused as 453 minimal-content and 58 empty extracts, with 295 examined individually — externally linked content and videos, videos containing sound but no narration, PDF-type nuggets referencing an unsupported document store, and image-only infographics. Treating AI output as provisional and evidencing its limits makes the results usable in a legal or audit context.



Figure 6 – Reports: General Insights - Content Extraction Analysis

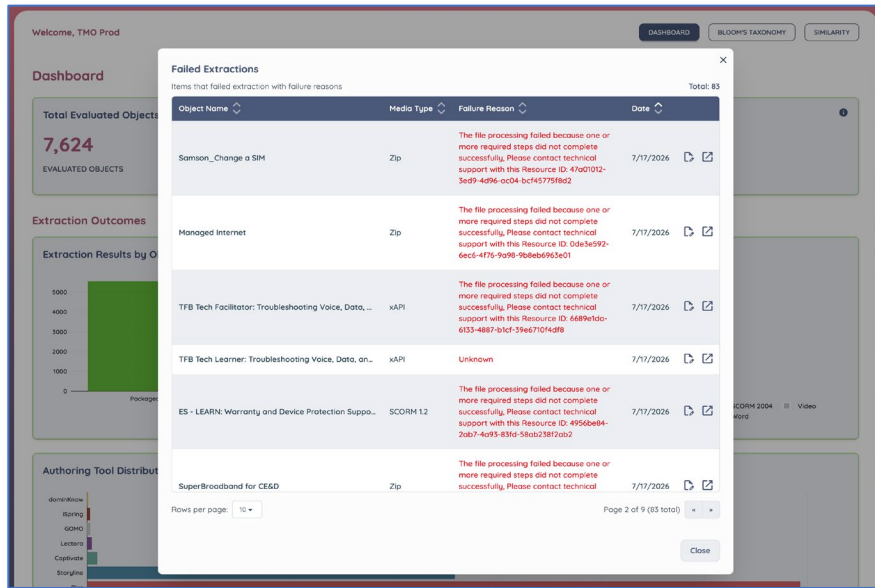


Figure 7 – Reports: General Dashboard - Failed Extraction Analysis

8. It is a continuous instrument, not a one-time audit

A content audit is an event whose findings decay immediately. MetaLark scans and indexes new content automatically on upload to the connected LMS or content management platform, so tagging, skills alignment and trend data stay current. Flagged and failed objects are reprocessed as the extraction pipeline improves. The corpus remains defined and understood as the library changes, which is what allows new content proposals to be assessed against known coverage indefinitely rather than for one budget cycle.

9. Overlap and duplication become visible and actionable

Because every object carries a description, accurate metatags and inferred skills, MetaLark can review the library for content-level similarity using inferred skills and thematic clusters. That surfaces clusters of duplicate and near-duplicate material, topics covered repeatedly by different business units, domains with concentrated coverage alongside domains with none, and content no longer matching any current business priority. It also produced a complete picture of authoring tool sprawl at the wireless carrier, driving a standardization decision to consolidate on Articulate Rise, Storyline and PDF. This is the capability that turns a content function from an order-taking operation into one that enters budget conversations with evidence.

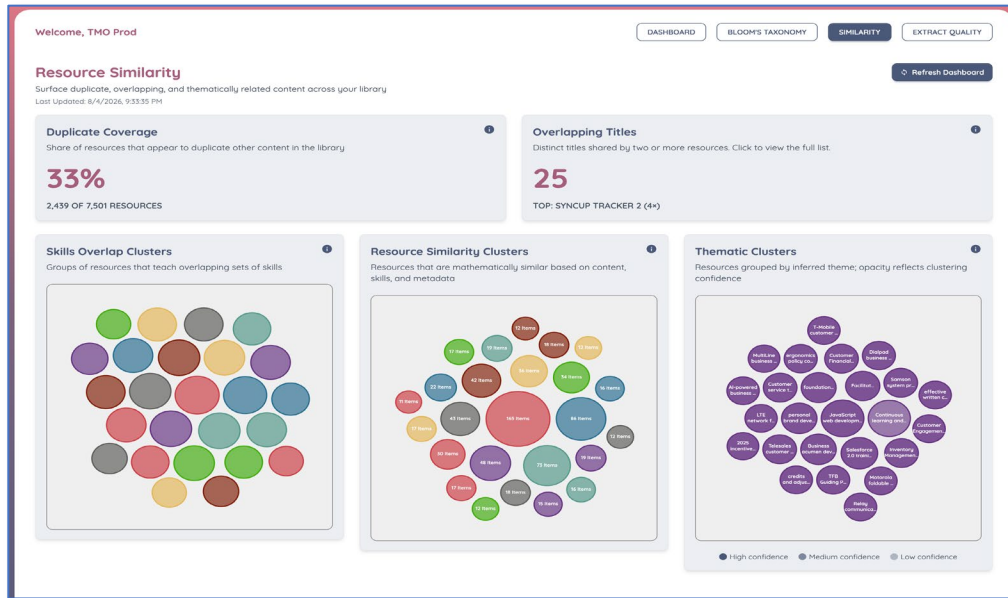


Figure 8 – Advanced Insights: Content Similarities & Patterns

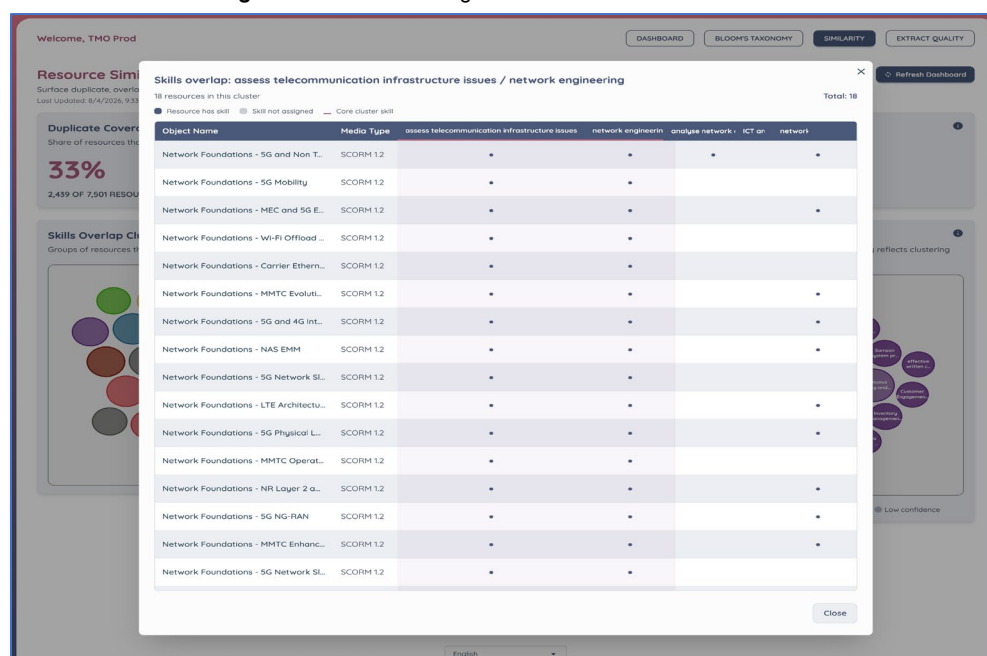


Figure 9 – Advanced Insights: Content Similarities & Patterns

0. Output is built for retrieval and agentic use

Extraction produces JSON and markdown optimized for analysis by a secure private language model, alongside vector-database indexing and content chunking. Agentic access is exposed through API and MCP. The practical effect is that the same extraction serves human search today and machine retrieval next, without a second pass over the library.

Technology Demo (based on judging criteria 4: Demo)

Access demo at:

<https://metalark.ai/bhg26-submission/>

The narrated walkthrough referenced above follows a single content library from ingestion through to strategic decision using a recorded demonstration.

- **Managed content view.** Content uploaded into MetaLark for extraction and analysis, showing the consolidated library across source systems and types.
- **Content detail.** An individual extracted resource with its AI-generated title and description, mapped skills and metatags, and a resource detail report for the selected object.
- **Content preview.** The same asset opened in the MetaLark player for immediate review without leaving the application.
- **Skills view.** Discovered and mapped skills summarized across the entire library.
- **Search.** Basic search with highlighted matches, advanced search with deep links that jump directly to the matched term inside the asset, and the embedded player launching the content positioned at the search results.
- **Reporting.** Content analysis by type, extraction analysis and failed-extraction diagnostics, and Bloom's taxonomy analysis at library and item level.
- **Advanced insights.** Content similarities and patterns (see appendix on mini-site) —the analysis supports consolidation, retirement and build-versus-reuse decisions.

Measurable Results (based on judging criteria 5: Measurable Results)

The results below are drawn from an actual production deployment at the wireless carrier, a Fortune 100 enterprise, documented in the MetaLark Content Extraction Report of 11 June 2026. The bulk extraction run executed on 29 May 2026 across the production LMS environment, with a delta catch-up rerun in mid-June 2026.

Measure	Count	Rate
Learning objects evaluated (in scope)	7,003	—
Successfully extracted	6,834	97.59%
Courseware packages extracted	4,970 of 5,101	97.43%

Measure	Count	Rate
Digital files extracted (PDFs, videos, images)	1,864 of 1,902	98.00%
Clean content extracts	6,323	92.52%
Content items flagged for human review	511	7.48%
Failed content extractions	169	2.41%

Scope note: the analysis covered active and inactive, published and unpublished courseware packages and digital files. Excluded by agreement with the client were objects with fewer than six assigned users, HTML/text and URL nuggets, foreign-source objects, shared replica objects, and objects with missing assets.

Extraction results — metadata and skills

Measure	Count	Basis
Unique metatags surfaced and discovered	31,309	6,323 objects
Unique skills mapped to content	1,901	6,323 objects
Content objects found with incomplete metadata	2,421 of 7,003	34.58%
Content objects with updated descriptions	7,003 of 7,003	100.00%

Scale and composition of the library processed

The composition of the library is itself a measurable result, because it establishes that the volume was not addressable by conventional tooling. Packaged courseware accounted for 4,959 of the objects, none of which was readable by any commercial search tool or language model prior to extraction.

Packaged content type	Successful extracts
SCORM 1.2	2,664
xAPI	1,926
SCORM 2004	369
Other (non-tracking HTML)	11

Digital resource type	Successful extracts
PDF	1,495
Video	291
HTML nugget linked to document store	78

Authoring tool	Successful extracts
Articulate Rise	3,109
Articulate Storyline	1,642
Generic HTML-based Content	137
Adobe Captivate	38
ELB Learning Lectora	18
iSpring	12
Gomo	11
dominKnow	3

Benefits and impact for the L&D function

- **Audit capability delivered in hours rather than months.** 7,003 objects were assessed and 6,834 indexed in an end-to-end content review completed in 36 hours of processing, against an estimated 1,750 person-hours — approximately eleven person-months — for equivalent manual review at 15 min/object.
- **Metadata coverage transformed.** 6,834 objects were updated with more than 31,000 AI-generated tags, descriptions and summaries, addressing a library in which 2,421 objects (34.58%) had incomplete or inaccurate metadata. Concept-level discoverability improved correspondingly.
- **Time to answer reduced from weeks to hours.** Content questions that previously took weeks now resolve in minutes or hours. In one instance, L&D leadership requested every occurrence and variation of a phrase from a legacy HR program deemed inappropriate and culturally misaligned with current program values; MetaLark returned exact + semantic results in < 60m, enabling fast action.
- **A new standing with Legal.** Remediation of flagged language became a triaged worklist with named owners and an audit trail, in place of an open-ended review. The L&D function delivered Legal something it had not previously been able to offer: a complete, defensible and re-runnable analysis. A working relationship now exists between the two functions because L&D proved to be the team that could answer the question.
- **Position in the enterprise talent architecture.** The function moved from supplying content lists with incomplete metadata to being the authoritative upstream source of vetted content and associated skills, feeding the separately selected skills intelligence platform that aligns earned skills to completed learning.
- **Content investment decisions made on evidence.** With every object carrying a description, accurate metatags and inferred skills, duplicate and near-duplicate clusters, repeated topic coverage across business units, over-served and unserved domains, and content no longer matching a business priority all became visible. Content requests are

now assessed against known coverage, supporting evidenced build-versus-reuse decisions and consolidation and retirement choices.

- **Authoring tool sprawl quantified and addressed.** Extraction produced a complete picture of tool sprawl across eight authoring tools, which is driving a standardization decision to consolidate on Articulate Rise/Storyline and PDFs.
- **Direct audit cost reduced.** The client has substantially reduced the direct expense of content audits while improving library health, freshness and consistency.
- **Designer and SME capacity returned.** Instructional designers and subject-matter experts recovered measurable time previously absorbed by file review exercises, redirected to new content design and delivery, SME and stakeholder engagement, and strategic planning. No client L&D role was reduced or eliminated as a result.

Benefits and experience for learners

- **Discoverability.** A streamlined, concise library returns more accurate and directed results to admin or future learner queries.
- **Currency.** Retiring outdated and duplicate content means enterprise language model and agentic search draw on more accurate and concise sources, so learners are no longer served superseded or non-compliant material.
- **Skills visibility.** Vetted skills are now tied to content completions, measured by the skills intelligence platform and surfaced through the LMS and an in-the-flow learning agent.

Quality measurement and disclosed limits

Extraction quality was measured and reported rather than assumed. Of 6,834 extracts created, 511 (7.48%) were flagged as needing review by AI-driven quality analysis, and 169 objects (2.41% of scope) failed extraction outright. Flagging identifies items warranting closer inspection and does not necessarily indicate a content problem.

Root cause of review flag	Count	Share of flagged
Minimal extractable content	453	88.65%
Empty extract	58	11.35%

Of the 511 flagged items, 295 were reviewed in depth and the patterns documented and reported to the client with a remediation plan and named owners.

Pattern identified	Count
Courses with externally linked video	100
PDF-type nugget linking to an unsupported document store	78
Externally linked content (dispatch)	55

Pattern identified	Count
Video with sound effects only, no spoken audio	41
PDF infographics with minimal text	21

Each pattern converted into scheduled engineering work with published status: image-based PDF OCR extraction and enhanced image handling entered an upcoming release, special-character handling in filenames and object names was corrected in the run itself, unsupported media file skipping was completed, and external content crawling remains under consideration. All items flagged for review and all failed extracts are queued for reprocessing once the enhanced image handling releases, which is expected to resolve a portion of the minimal-content and empty-extract cases.

Governance and responsible practice

- **AI deferred to human judgment.** AI-generated metatags were applied only where no manually assigned tag already existed, by explicit agreement with the customer. Nothing authored by an L&D resource was overwritten by AI inference except where specifically directed and no content was retired, altered or republished; disposition recommendations were flagged/reviewed by named human owners.
- **Data remained governed.** All processing ran in the client's dedicated LMS environment under SOC 2 controls with enterprise SSO. No client content was used to train external AI models, under strict data residency and retention terms.
- **The human consequence was the subtraction of drudgery, not of roles.** The alternative was never no work; it was approximately 1,750 person-hours of manual file review absorbed by instructional designers and subject-matter experts. Reclaimed capacity was reapplied to content design and development, stakeholder engagement and strategic planning.

Statement of maturity

MetaLark.ai services went live at the wireless carrier in early June 2026. The extraction layer is fully enabled and in continuous use across all content uploaded to the client's LMS and document management solution, so the corpus stays defined and current as content changes rather than decaying after a single audit. Review and analysis practices continue to be documented, with scheduled reviews of flagged and failed objects and periodic audits of metadata and inferred skills. With ingestion proven, the client's L&D team is investigating learner-facing concept-level retrieval through the MetaLark content player and an enterprise agentic interface, using the vector database and content chunking for pinpoint delivery rather than object-level launch.