

Is Your Content AI-Ready?



A Content Hygiene Checklist for Associations

1. PDF & Document Quality

- Documents are text-based, not scanned images— *text should be selectable and copy-able*
 - Long PDFs are broken into logical sections or separate articles where possible
 - PDFs avoid mixing editorial content with ads or unrelated material
 - Publications have a text-based equivalent (HTML, DOCX, or text-based PDF)
 - Scanned or legacy documents that need OCR have been identified, flagged, and actioned
-

2. File Organization & Version Control

- Files follow a consistent naming convention
Example: 2024-Annual-Report-FINAL.pdf – not Report_v7_FINAL_v3_use-this-one.pdf
 - Outdated versions are archived or removed, not left alongside current files
 - Documents include clear dates (publication date, revision date, or both) in document metadata
 - There is a single authoritative location for each document type
 - Duplicate files across platforms (website, SharePoint, Google Drive) have been identified
-

3. Website Content Structure

- Content follows a logical hierarchy reflected in URL structure
Example: /resources/reports/2024-annual-report – not /page?id=4827
- PDFs have dedicated landing pages with titles, descriptions, and dates – *A landing page makes the content findable in search engines and AI systems – even if the full document is member-only.*
- Consistent taxonomy and categories are used across content sections
- Broken links to PDFs or documents have been identified and fixed
- Member-only content is clearly gated – not accidentally public or accidentally hidden

4. Content Freshness & Governance

- There is a defined process for archiving outdated content and flagging content that is not evergreen for future archiving
 - Someone owns the decision of when content should be updated, retired, or replaced
 - Community-generated or unverified content is clearly distinguished from authoritative content
 - AI or AI-assisted content is reviewed for accuracy before publishing
 - A content audit has been done in the last 12–18 months (or is planned)
-

5. Video & Audio Content

- All video and audio content has accurate text transcripts - *Auto-generated captions are a start but should be reviewed for accuracy – especially for technical or industry-specific language.*
 - YouTube content is organized in playlists with captions enabled
 - Vimeo content has auto-transcription enabled (or transcripts are provided separately)
 - Webinar recordings include transcripts or summary documents
-

6. DISCOVERABILITY & METADATA

- Documents and pages include descriptive titles, not just file names
- Key content has a short summary or abstract that can surface in search results
- Documents that live publicly online include links to related or source content
- Microsites, subdomains, event sites, and learning portals have been inventoried - *AI tools and crawlers need to know these exist. If they aren't disclosed or linked, the content may be missed entirely.*

Where do you go from here? You don't need to fix everything at once! Knowing where the gaps are is the first step. Focus on one section at a time and reach out to Betty if you need help prioritizing!



meetbetty.ai