Who Speaks for the Organisation

Owning your AI narrative

The organisation is now judged by content it may no longer fully control. Years of published material shape reputation, competitive standing, and regulatory exposure, while AI systems now read, summarise, and represent that content at scale.

> *"The board implication is also clear:* ***unmanaged content is no longer passive archive material,*** *because it can now influence how AI systems explain the organisation to the market."*

In summary

The content problem you do not want is **your AI advantage**

Introduction

AI is a new reader of what the organisation publishes. It sits alongside the customers, partners, regulators, and people the organisation has always served.

This paper is about that ground: how the organisation is represented when the reader is no longer only human, and what sits inside reputation, competitive standing, and the regulatory position when the same content is reaching both audiences.

Competitive position is part of that ground. The narrative being read about the organisation is being read by AI alongside everyone else, and the question of whose narrative it is, and who has assembled it, sits inside the picture the paper covers.

The regulatory position carries two strands.

1. The first is the accuracy of the information being delivered, including historic material still online that no longer represents the organisation's current position.
2. Second is the obligation, both reputational and legal, to ensure the content is inclusive and accessible to the people relying on it.

Executive summary

Future visibility, competitiveness, and how reputation is managed are the concerns the board and the executive carry. AI (LLM/Agents) is a new reader of the content the organisation publishes, sitting alongside the customers, partners, regulators, and people the organisation has always served, and the same content is now reaching both.

The assets the organisation relies on, annual reports, statements, prospectuses, prescribing information, brochures, and press releases, are published, republished, and circulated across the organisation's digital properties year after year. They are the material the organisation is read by, with many published as PDFs.

The legal position on this material is developing across the US, the UK, and the European Union, although the route and scope differ by jurisdiction and organisation type.

1. Department of Justice ADA Title II compliance dates 20 April 2027 (US Department of Justice). Title III, state-level digital accessibility laws, and the HHS Section 504 rule continue to apply (Jackson Lewis, 2026).
2. European Accessibility Act applies from 28 June 2025 (European Commission, 2019),
3. UK Public Sector Bodies Accessibility Regulations 2018 and the Equality Act 2010 remain in force (UK Government, 2010, 2018).

Brands invest in the visual and editorial care of their published material. That care is most visible in brochures, prospectuses, and the assets that carry brand identity. Academic research has measured the gap between how this material is read by people and how it is read by AI (Koreeda and Manning, 2021; Volpini et al., 2026).

> Example.
>
> *Job applications published as PDFs years ago can remain online and visible. They reference roles and entities that no longer exist, and the terms in them may no longer be compliant with current employment legislation. Third parties carry copies, and the material is read alongside current content when AI assembles a picture of the organisation.*

McKinsey research finds that brand-owned sites comprise only 5 to 10% of the sources AI search references in many categories (McKinsey, 2025). The other 90 to 95% comes from publishers, affiliates, user-generated content, and third-party material. Material the organisation has published over many years remains visible in that wider picture.

The picture AI assembles is the picture the market acts on.

Commercial obligation: inclusion in the AI sense

The first obligation has arrived on the same content. AI agents, retrieval systems, and answer engines now read what the website holds and describe the organisation to the market on the basis of what they find.

AI mediates discovery

Google has expanded AI Overviews to more than 200 countries and territories and to more than 40 languages, with usage growth above 10% in major markets where AI Overviews have been deployed (Google, 2025). Google's developer guidance for succeeding in AI search now explicitly addresses how content should perform in AI search experiences, not just traditional search results (Google Search Central, 2025).

OpenAI (2026). Powering product discovery in ChatGPT, with structured merchant feeds governing discoverability, relevance, and accuracy (OpenAI, 2025). The merchant guidance is explicit that structured product feeds materially improve how products are found, compared, and presented to the user.

Adobe reports that traffic from AI sources to US retail sites grew 393% year on year in Q1 2026, and that AI-sourced traffic converted 42% better than other traffic in March 2026 (Adobe, 2026). Salesforce reports that AI and agents accounted for $262 billion of 2025 holiday spend (Salesforce, 2025).

Deloitte's 2026 Global Retail Industry Outlook reports that more than 9 in 10 retail leaders expect AI to be used more than traditional search by 2026, and that 50% expect today's multi-step shopping journey to collapse by 2027 into a single AI-driven interaction (Deloitte, 2026).

Harvard Business Review reports that in a 2025 survey of 12,000 consumers, 58% had turned to generative AI tools for product or service recommendations, against 25% in 2023 (Dubois, Dawson and Jaiswal, 2025). The same article reports a 1,300% surge in AI search referrals to US retail sites during the 2024 holiday season.

BCG's 2025 Black Friday Consumer Study, surveying 10,240 consumers across 10 countries, reports that 48% had used or planned to use generative AI during the year-end sales events, a 9 percentage-point increase over 2024 (BCG, 2025). The same study confirms the shift across markets and product categories.

Measurement is breaking down

The metrics that defined search success for the past two decades are no longer measuring what matters. Forrester reports that clicks were the primary measure of both paid and organic search performance (Forrester, Lai et al, 2025). In an AI-mediated environment, organic success is being redefined as frequent appearance in AI-assisted conversations, and the established measurement framework is moving with it.

McKinsey research finds that just 16% of brands today systematically track AI search performance (McKinsey, 2025). The other 84% are operating without a measurement framework against the channel that increasingly mediates discovery for them.

Similarweb reports that zero-click search on Google grew from 56% to 69% in the year following the launch of AI Overviews (Similarweb, 2025). The audience has not disappeared. The attention is being intercepted before it reaches the source.

Authoritas research, cited by the Professional Publishers Association, reports that AI Overviews reduced click-through rates by up to 57%, accelerating the need for publishers to diversify how audiences discover their content (Authoritas / PPA, 2025). DMG Media reported to the UK Competition and Markets Authority that click-through rates on certain search terms dropped by 89% on desktop and 87% on mobile when AI Overviews appeared (DMG Media, 2025).

Enders Analysis, also cited by the Professional Publishers Association, reports that half of UK media groups recorded a decline in search traffic across 2025 (Enders Analysis / PPA, 2025). The shift is not isolated to a single publisher or category.

Trust and provenance

The US National Institute of Standards and Technology Generative AI Profile identifies provenance harms, including misinformation, disinformation, deepfakes, and tampered content, as risks requiring active identification, ranking, and mitigation (NIST, 2024). Provenance and information integrity are now part of practical AI risk management.

The UK Information Commissioner's Office has issued current guidance on the use of storage and access technologies, covering tracking pixels, link decoration, navigational tracking, device fingerprinting, web storage, and scripts and tags (Information Commissioner's Office, 2025). Privacy and consent integrity are part of how an organisation is read by automated systems.

How AI reads documents

AI systems do not read documents the way a human reader does. They operate on retrieval-augmented generation, where the language model is grounded in retrieved passages from external sources. The accuracy of what the model produces depends on the accuracy of what the retrieval system can find and parse.

i. Lin (ChatDOC, 2024) demonstrates empirically that PDF parsing accuracy materially affects retrieval-augmented generation quality. Where PDF structure recognition is improved, retrieval accuracy improves with it. Where it is not, the language model has weaker source material to ground its answers in.

ii. Hui, Lu and Zhang (Tsinghua University and National University of Singapore, 2024) introduced UDA, a benchmark suite of 2,965 real-world documents and 29,590 expert-annotated question-answer pairs, peer-reviewed and published at the NeurIPS Datasets and Benchmarks Track. The benchmark demonstrates that parsing unstructured documents is error-prone, and that document structure materially affects retrieval performance.

iii. Volpini, Raad, Gamba and Riccitelli (WordLift, 2026) report that enhanced entity pages with structured linked data produced a 29.6% accuracy improvement in standard retrieval-augmented generation and a 29.8% accuracy improvement in agentic retrieval, across 2,443 query evaluations in 4 industry verticals. Structure measurably affects what AI retrieves and how accurately it answers.

GEO sits inside this picture

Generative Engine Optimisation (GEO) describes the practice of structuring content so AI systems can read, interpret, and surface it accurately. Aggarwal et al. (Princeton University and Georgia Institute of Technology, KDD 2024) introduced the term in the academic record.

GEO sits inside the AI obligation, not alongside it. Without the structural inputs the discipline depends on, GEO effort produces limited results regardless of how well it is executed. The structural inputs are the same inputs the legal obligation depends on.

Legal obligation: accessibility in the human sense

The legal obligation on the PDFs linked from a public-facing website is settled in law. It applies across the US, the UK, and the European Union, and the direction across each jurisdiction is consistent: a planned, evidenced, improving approach is what the regulations contemplate.

NOTE
The legal obligation on PDFs linked from public-facing websites is increasingly clear, although the precise duty depends on the organisation, jurisdiction, and service involved.

United States

The Department of Justice issued the Title II final rule on 24 April 2024, requiring state and local public entities to ensure their web content and mobile applications meet WCAG 2.1 Level AA standards. Compliance dates were originally set at 24 April 2026 for entities serving 50,000 people or more, and 26 April 2027 for smaller entities and special district governments.

On 20 April 2026, the Department of Justice published an Interim Final Rule extending the compliance dates by 1 year (US Department of Justice, 2026; 91 Fed. Reg. 20902). The compliance date for entities serving 50,000 or more moved to 26 April 2027. The compliance date for entities under 50,000 and special district governments moved to 26 April 2028. The Interim Final Rule took effect on the day of publication. The Department stated it fully anticipates implementing the regulation at the new deadline.

The extension does not suspend the underlying non-discrimination and effective-communication obligations of Title II, which have long supported web accessibility claims without a specific technical standard (Duane Morris, 2026).

Private plaintiffs and advocacy organisations retain the ability to bring claims during the extension period.

Title III of the ADA, covering private businesses operating as places of public accommodation, remains in force. State-level digital accessibility laws continue to apply across multiple US jurisdictions and are enforced independently of the federal Title II calendar. The Department of Health, theHuman Services Section 504 rule, which imposes parallel web and mobile accessibility requirements on recipients of HHS funding, retained its 11 May 2026 first compliance date and was not extended (Jackson Lewis, 2026).

Section508.gov directs federal agencies and contractors to publish content as accessible HTML rather than PDF, and treats PDF as the format to avoid or replace where alternative formats are available (Section508.gov, 2024).

United Kingdom

The Public Sector Bodies (Websites and Mobile Applications) (No. 2) Accessibility Regulations 2018 require public sector bodies to make their websites and mobile applications accessible, to publish an accessibility statement, and to document any reliance on the disproportionate burden provision through a formal assessment (UK Government, 2018).

The Equality Act 2010 creates a continuing anticipatory duty to make reasonable adjustments. The duty is anticipatory: organisations are expected to consider in advance what adjustments users with disabilities may need, rather than waiting for a specific individual request (UK Government, 2010).

The Government Digital Service published clear guidance in 2018 that content should be published in HTML rather than PDF wherever possible, on the grounds that PDFs are not designed for reading on screens and create barriers that HTML does not (Government Digital Service, 2018). The guidance has not been withdrawn or softened.

European Union

The European Accessibility Act (Directive (EU) 2019/882) applies from 28 June 2025 across a defined set of products and services offered in the EU market. Its scope is specific: e-commerce, banking, e-books, transport ticketing, smartphones and computers, ATMs, and a defined set of related services (European Commission, 2019).

Where the EAA applies, it includes a concept of disproportionate burden that requires evidence, a documented assessment, and active demonstration that the organisation is working to reduce barriers over time. An organisation that cannot show a plan, evidence of progress, and a structured approach is not in a position to rely on disproportionate burden as a defence.

The direction of travel

Across the US, the UK, and the European Union, the direction of travel is consistent. HTML is the preferred publishing format. PDF is the format to replace where an alternative is available. A planned, phased, improving approach is the model the regulations reward.

An organisation that knows what its PDF estate contains, that has a plan for converting it into inclusive web content, and that can show evidence of progress sits inside the model the regulations contemplate. Without that knowledge, the organisation cannot answer the question of what its estate contains, and cannot demonstrate the planned approach the regulations describe.

Why PDFs are the pivot point

Inclusion is no longer only about the people in front of the screen; it is also about the AI systems behind it, and PDFs are where both obligations either meet or fail together.

The structural cause

PDFs were designed as a print format. Adobe's Portable Document Format was built to preserve visual presentation across devices and operating systems, not to encode semantic structure that machines can interpret reliably. The structure a reader sees on screen, headings, paragraphs, lists, tables, reading order, is rendered visually rather than encoded as data.

Rahman and Finin demonstrate that PDF rendering tools are not designed to support post-processing. Different valid ways of producing the same visual result mean that no structure can reliably be derived from how the text operators are used (Rahman and Finin, 2017). The visual result and the underlying structure are not the same thing.

Koreeda and Manning (Stanford University, 2021) show that conventional PDF-to-text tools fail at fine-grained logical structure extraction. Their multimodal transition parser produced a paragraph boundary detection F1 score of 0.951 against a popular PDF-to-text tool's 0.739. The gap measures the structural problem directly.

The structural barrier that excludes a person using assistive technology is the same structural barrier that excludes the AI system describing the organisation. The screen reader cannot derive the reading order. The retrieval system cannot derive the semantic structure. A single structural cause produces both consequences on the same documents.

The convergence is named in the academic record

The accessibility literature and the machine-interpretability literature have been treating the same structural problem from 2 angles. The convergence is now being named directly. Walsh, Boudreaux et al. report on using AI to convert PDF documents to machine-readable HTML with semantic tags, extracted figures and tables, and generated summaries, with the explicit dual aim of making them machine-readable and more accessible to a wider audience (Walsh, Boudreaux et al., 2023).

A benchmark dataset of scholarly PDFs with expert-validated accessibility annotations across 7 WCAG and PDF/UA criteria, including alternative text quality, logical reading order, semantic tagging, and table structure, has been published with an LLM-based framework for accessibility assessment (Benchmarking PDF Accessibility Evaluation, 2025). The same structural elements that determine accessibility for the screen reader user are the structural elements that determine machine interpretability for the AI.

Iliadis (Journal of the Association for Information Science and Technology, 2023) sets out how Schema.org and structured markup operate in search and knowledge discovery, and the role of structured data infrastructure in retrieval. Semantic markup affects machine readability. The accessibility literature and the information retrieval literature converge on the same structural concept.

What this means for the PDF estate

The PDF estate is the operational point where the 2 obligations meet. A document that fails the screen reader is also more likely to be poorly parsed, misread, or underused by AI systems describing the organisation.. A document that the AI cannot interpret is a document the comparison set built by AI does not carry. A document the comparison set does not carry is a document the customer never sees.

Treating the 2 halves as separate workstreams produces double the cost and half the coverage. The structural fix that addresses the legal obligation is the structural fix that addresses the commercial one, applied to the same files in the same operation.

The design-led PDF problem

Approximately 26% of PDFs are produced as design-led pieces, where the visual experience is the point and structure are not built in. These documents are the hardest case for AI interpretation, and the consequence is felt sharpest at the high end of the market.

> ***….documents built to be looked at - visually enjoyed, not read by a machine.***

What a design-led PDF is

A design-led PDF is a document where the visual presentation carries the meaning. Brochures for super-yacht builders, lookbooks for designer fashion houses, sales documents for luxury vehicles, catalogues for premium watch brands, tasting notes for fine wines and spirits, auction guides for art and auction houses, prospectuses for private banking, brochures for luxury property and super-prime developments.

The visual experience is the point of these documents. Composition, typography, photography, white space, and layout are doing the work of communicating the brand. Structure that a machine can interpret is not built in, because the document was built for a different purpose.

The Hui, Lu and Zhang benchmark (NeurIPS 2024) demonstrated measurably that unstructured documents pose specific retrieval challenges. The Volpini et al. study reports a 29.6% retrieval accuracy gap between plain documents and documents with usable structure. The design-led PDF is the case at the far end of that gap.

Why this is felt at high-end brands

AI agents now mediate the comparison set the buyer sees before any human contact. A buyer asking AI to compare super-yachts, designer watches, or luxury property options is presented with a comparison built from what AI has been able to read.

A brand whose product information lives in a design-led PDF is a brand AI cannot read at the level required to populate the comparison. The comparison is built from the brands AI could read. The brand that did the most distinctive design work on the PDF is the brand the comparison leaves out.

The consequence is structural. The brand has not been judged. It has not been compared and rejected. It has not appeared in the comparison at all, and the buyer has moved on to the brands the AI could describe, with no awareness that the missing brand was an option.

The compounding effect

Design-led PDFs accumulate over time. New collections, new vintages, new model years, new property launches, new collaborations. Each is published as a new document, designed to the same standard, structured the same way.

The estate of design-led documents grows, and the AI-readable representation of the brand stays where it was.

The brand that does the most design work is the brand most affected. The investment in the document is what makes the document hard for AI to read, and the same structural fix applies in each case.

The unknown estate and stale messages

The unknown estate compounds the problem. Sitemorse data indicates 41% of websites are unknown to the organisation's digital teams (P&C and Sitemorse risk profiling 2017 to 2023, covering 100+ million websites). The same internal analysis indicates 19% of PDFs on organisational websites are duplicates.

The unknown estate is the part of the digital footprint still visible externally but no longer actively governed. Old domains, archived campaigns, third-party hosted material, forgotten subdomains, obsolete reference pages, public material created for earlier operational needs.

What the unknown estate carries

Out-of-date offers continue to circulate. Wrong product names persist. Term issues, superseded guidance, and stale claims continue to feed the picture AI forms of the organisation. AI does not respect internal organisational memory. What is online is what AI uses, regardless of whether the organisation still considers it current.

Our data indicates 37% of PDFs become untracked once they leave the producer's website, including through circulation in annual reports and product catalogues (data). Mirrored copies on partner sites, supplier portals, sector trackers, and cached search indexes carry whatever was downloaded at whatever point. A correction at source does not reach the mirrors.

Most of the AI picture sits outside owned content

McKinsey research shows that brand-owned sites comprise only 5 to 10% of the sources AI search references in many categories (McKinsey, 2025).

The other 90 to 95% comes from publishers, affiliates, user-generated content, third-party sites, and external references the organisation does not own.

AirOps research analysing more than 21,000 brands found that 85% of brand mentions in AI search come from third-party content, not the brand's own site (AirOps, 2025). Owned content is a baseline. The picture AI carries is built across the wider web, and the wider web includes whatever historic material the organisation still has visible.

Why this matters for the AI obligation

AI systems describe the organisation on the basis of what they read across the public internet, not on the basis of what the organisation considers current. A page the organisation has stopped looking at is a page AI is still using to describe the organisation. A document the organisation no longer publishes is a document AI is still presenting as the organisation's position.

The picture AI forms is the picture the organisation has stopped managing. The cost is the gap between what the organisation says about itself today and what AI says about the organisation tomorrow.

Where automation handles the work

Automation reads the design-led PDF, recognises structure that was never built in, and produces inclusive web content the screen reader user, the AI engine, and the AI agent can each use. AAAnow / Sitemorse internal data indicates an 85% improvement in accessibility outcomes from automated HTML conversion against the established manual baseline (AAAnow / Sitemorse internal data).

What the work does

Sites are located. PDFs across the estate are identified through automated discovery, including those hosted on external servers. Each PDF is converted into structured, inclusive web content, and the original PDF is preserved alongside the converted content.

The same conversion handles all 3 audiences.

1. The screen reader user gets content with logical reading order, semantic structure, and reflowing layout.
2. The AI engine gets content with usable structural markup.
3. The AI agent gets content it can read, summarise, and present in the comparison set on the buyer's behalf.

The work that was previously a project is converted into an ongoing service. New PDFs are identified and converted automatically as they are published. The estate stays current without effort from internal team s.

Why AI agents matter most of the 3

Of the 3 audiences served by a single conversion, AI agents matter most. They increasingly influence which sites are found, compared, shortlisted, and chosen on the buyer's behalf. The buyer asking the agent to find, compare, recommend, or shortlist is presented with a result built from the sites the agent could read.

The brand AI agents can read is more likely to be represented accurately, compared fairly, and recommended. Where AI agents cannot read the brand, the brand does not appear in the recommendation, the structural fix that produces the inclusive web content is the structural fix that gets the brand into the comparison set.

Implications for the board

The implication for the board is direct. AI agents now decide which sites get chosen on the buyer's behalf, and the picture they carry is the picture the market acts on. The board either owns that picture, or accepts the one AI assembles from what is currently online.

Control of the narrative

Reputation is now a function of what AI describes, ranks, and recommends on the basis of what is published. Harvard Business Review reports that AI is shifting the buyer relationship in two ways at once: conversational AI displacing websites and traditional search as the way consumers learn about products, and AI agents acting as buyers, making purchasing decisions on behalf of humans (Puntoni, 2026). Both shifts favour brands whose information is structured, trusted, and easy for AI systems to synthesise.

MIT Sloan Management Review reports that even market-leading brands risk becoming invisible if they continue with familiar SEO practices, and that companies adapt by reconsidering how they prioritise and allocate resources for AI-driven search (Pettiette and Whitler, 2026). The organisation that owns its picture does the work that lets AI read it accurately.

With the work done, the organisation is read by AI as it has set itself up to be read. AI uses the organisation, describes it accurately, and represents it in comparison environments. The organisation is positioned to be cited and shortlisted by AI agents acting on the buyer's behalf.

The cost of waiting

AI is not waiting for the organisation to be ready. It is using what is online now. Once a picture is formed and reused, it is the picture that gets surfaced, summarised, and acted on.

The cost of waiting is the cost of letting the external picture set without intervention. Competitors that move first are read accurately by AI first. The picture they own becomes the picture the comparison set is built around. The organisation that acts later is correcting a picture that has already formed, against competitors AI is already representing.

What the board does next

The work begins with the question the rest of the paper has answered. The PDF estate sits inside 2 obligations at once. The structural fix that addresses one is the structural fix that addresses the other. A single conversion serves the screen reader user, the AI engine, and the AI agent.

The board reviews the position, sets the target, and commissions the work.

Conclusion

The PDF estate now sits between the organisation and the people, the systems, and the agents that read it. The legal obligation is real, settled in law, and the organisation's to act on. The commercial obligation has arrived on the same content, and AI is using it now.

Inclusion is no longer only about the people in front of the screen; it is also about the AI systems behind it, and PDFs are where both obligations either meet or fail together.

The board owns the picture AI forms of the organisation, or accepts the one AI assembles from what is currently online. The work that lets AI read the organisation accurately is the same work that meets the legal obligation, applied to the same files.

References

1. Adobe (2026). AI traffic surge to retail sites. Q1 2026 retail data. Adobe Digital Insights. https://business.adobe.com/blog/ai-traffic-surge-retail-sites-not-machine-readable
2. Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., and Deshpande, A. (2024). GEO: Generative Engine Optimization. Princeton University and Georgia Institute of Technology. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2024). https://dl.acm.org/doi/10.1145/3637528.3671900
3. AirOps (2025). The Influence of Offsite Signals in AI Search. Research analysing more than 21,000 brands. Authored by Alex Halliday and team. https://www.airops.com/report/the-influence-of-offsite-signals-in-ai-search
4. Authoritas / PPA (2025). Click-through rate impact of AI Overviews. Cited by Sajeeda Merali, CEO of the Professional Publishers Association. https://www.inpublishing.co.uk/articles/the-year-things-changed-again-26248
5. BCG (2025). Black Friday Consumer Study, October 2025. Survey of 10,240 consumers across 10 countries. Boston Consulting Group. https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap
6. Benchmarking PDF Accessibility Evaluation (2025). A Dataset and Framework for Assessing Automated and LLM-Based Approaches for Accessibility Testing. https://arxiv.org/abs/2509.18965
7. Cluttered to Clear (2025). Improving the Web Accessibility Design for Screen Reader Users in E-commerce With Generative AI. https://arxiv.org/abs/2502.18701
8. Deloitte (2026). 2026 Global Retail Industry Outlook. Deloitte Insights. https://www.deloitte.com/mt/en/Industries/consumer/perspectives/global-retail-industry-outlook.html
9. DMG Media (2025). Submission to the UK Competition and Markets Authority on AI Overviews and click-through rate impact. Cited via Search Engine Journal. https://www.searchenginejournal.com/impact-of-ai-overviews-how-publishers-need-to-adapt/556843/
10. Duane Morris LLP (2026). DOJ Extends ADA Title II Digital Accessibility Deadlines by One Year. https://www.duanemorris.com/alerts/doj_extends_ada_title_ii_digital_accessibility_deadlines_one_year_0426.html
11. Dubois, D., Dawson, J. and Jaiswal, A. (2025). Forget What You Know About SEO. Here's How to Optimize Your Brand for LLMs. Harvard Business Review, June 2025. https://hbr.org/2025/06/forget-what-you-know-about-seo-heres-how-to-optimize-your-brand-for-llms
12. Enders Analysis / PPA (2025). Research on UK media groups and search traffic decline across 2025. Cited by the Professional Publishers Association. https://pressgazette.co.uk/media-audience-and-business-data/uk-and-us-publishers-says-google-ai-is-harming-website-traffic/
13. European Commission (2019). European Accessibility Act, Directive (EU) 2019/882. Application from 28 June 2025. https://eur-lex.europa.eu/eli/dir/2019/882/oj
14. Forrester (Lai, N., Pfeiffer, E., Buten, J., Curran, R., Gately, L., Irion, R., Johnson, J., and contributors) (2025). GenAI Forever Changes All Forms Of Search. 24 March 2025. https://www.forrester.com/report/genai-forever-changes-all-forms-of-search/RES182189
15. Government Digital Service (2018). Why GOV.UK content should be published in HTML and not PDF. https://gds.blog.gov.uk/2018/07/16/why-gov-uk-content-should-be-published-in-html-and-not-pdf/
16. Google (2025). AI Overviews expansion and search update. https://blog.google/products-and-platforms/products/search/ai-overview-expansion-may-2025-update/
17. Google Search Central (2025). Succeeding in AI search. https://developers.google.com/search/blog/2025/05/succeeding-in-ai-search
18. Hui, Y., Lu, Y., and Zhang, H. (2024). UDA: A Benchmark Suite for Retrieval Augmented Generation in Real-world Document Analysis. Tsinghua University and National University of Singapore. NeurIPS 2024 Datasets and Benchmarks Track. https://arxiv.org/abs/2406.15187

19. Iliadis, A. (2023). One schema to rule them all: How Schema.org models the world of search. Journal of the Association for Information Science and Technology. Wiley. https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24744
20. Information Commissioner's Office (2025). Guidance on the use of storage and access technologies. https://ico.org.uk/for-organisations/direct-marketing-and-privacy-and-electronic-communications/guidance-on-the-use-of-storage-and-access-technologies/
21. Jackson Lewis P.C. (2026). DOJ Extends Public Entities' Compliance Deadline for ADA-Related Website Accessibility, But HHS's May 2026 Deadline Still Looms. https://www.jacksonlewis.com/insights/doj-extends-public-entities-compliance-deadline-ada-related-website-accessibility-hhss-may-2026-deadline-still-looms
22. Koreeda, Y., and Manning, C. D. (2021). Capturing Logical Structure of Visually Structured Documents with Multimodal Transition Parser. Hitachi America and Stanford University. https://arxiv.org/abs/2105.00150
23. Lin, D. (2024). Revolutionizing Retrieval-Augmented Generation with Enhanced PDF Structure Recognition. ChatDOC. https://arxiv.org/abs/2401.12599
24. McKinsey (Silliman, E., Boudet, J., Robinson, K., Oppong, D., and Shah, N.) (2025). New front door to the internet: Winning in the age of AI search. McKinsey Growth, Marketing & Sales Practice. 16 October 2025. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/new-front-door-to-the-internet-winning-in-the-age-of-ai-search
25. NIST (2024). Artificial Intelligence Risk Management Framework: Generative AI Profile (NIST AI 600-1). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
26. OpenAI (2025). Powering product discovery in ChatGPT. https://openai.com/index/powering-product-discovery-in-chatgpt/
27. Pettiette, M., and Whitler, K. A. (2026). Can Customers Find Your Brand? Marketing Strategies for AI-Driven Search. MIT Sloan Management Review, 28 January 2026. https://sloanreview.mit.edu/article/can-customers-find-your-brand-marketing-strategies-for-ai-driven-search/
28. Puntoni, S. (2026). AI Is Upending Marketing on Two Fronts. Harvard Business Review, February 2026. Stefano Puntoni, Wharton Human-AI Research, University of Pennsylvania. https://hbr.org/2026/02/ai-is-upending-marketing-on-two-fronts
29. Rahman, M. M., and Finin, T. (2017). Understanding the Logical and Semantic Structure of Large Documents. https://arxiv.org/abs/1709.00770
30. Salesforce (2025). 2025 holiday shopping data. https://www.salesforce.com/uk/news/stories/2025-holiday-shopping-data/
31. Section508.gov (2024). Federal guidance on PDFs and accessible content. https://www.section508.gov/create/pdfs/
32. Similarweb (Barenholtz, L.) (2025). Zero-Click Searches: research showing zero-click search on Google grew from 56% to 69% in the year following the launch of AI Overviews. https://www.similarweb.com/blog/marketing/seo/zero-click-searches/
33. UK Government (2010). Equality Act 2010. https://www.legislation.gov.uk/ukpga/2010/15/contents
34. UK Government (2018). Public Sector Bodies (Websites and Mobile Applications) (No. 2) Accessibility Regulations 2018. https://www.legislation.gov.uk/uksi/2018/952/contents/made
35. US Department of Justice (2024). Title II of the Americans with Disabilities Act, Final Rule on Web Content and Mobile Application Accessibility. https://www.ada.gov/resources/web-rule-first-steps/
36. US Department of Justice (2026). Extension of Compliance Dates for Nondiscrimination on the Basis of Disability; Accessibility of Web Information and Services of State and Local Government Entities. Interim Final Rule, 91 Fed. Reg. 20902, 20 April 2026. https://www.federalregister.gov/documents/2026/04/20/2026-07663

37. Volpini, A., Raad, E., Gamba, B., and Riccitelli, D. (2026). Structured Linked Data as a Memory Layer for Agent-Orchestrated Retrieval. WordLift. https://arxiv.org/abs/2603.10700
38. Walsh, J., Boudreaux et al. (2023). Maximizing Equitable Reach and Accessibility of ETDs. https://arxiv.org/abs/2310.18427
39. AAAnow / Sitemorse internal data. P&C and Sitemorse risk profiling 2017 to 2023, covering 100+ million websites. Source of the 41% unknown estate, 19% PDF duplicate, 37% untracked PDF, 85% accessibility improvement, and approximately 26% design-led PDF figures used in this paper. Each labelled at the point of use. Internal AAAnow / Sitemorse data is not externally published. https://www.aaanow.com

Inventory of sources

Each source cited in this paper, with what it is and the link, listed in alphabetical order by author or organisation. Each source on the list also appears in the references above.

1. **AAAnow / Sitemorse data**

   AAAnow / Sitemorse data drawn from P&C and Sitemorse risk profiling 2017 to 2023, covering 100+ million websites. Source of the 41% unknown estate figure, the 19% PDF duplicate figure, the 37% untracked PDF figure, the 85% accessibility improvement outcome from automated HTML conversion, and the approximately 26% design-led PDF estimate.

   https://www.aaanow.com

2. **Adobe (2026)**

   Adobe Digital Insights data on AI traffic to US retail sites in Q1 2026. Source of the 393% year-on-year growth figure and the 42% better conversion figure for AI-sourced traffic.

   https://business.adobe.com/blog/ai-traffic-surge-retail-sites-not-machine-readable

3. **Aggarwal et al. (KDD 2024)**

   Princeton University and Georgia Institute of Technology peer-reviewed conference paper introducing the term Generative Engine Optimization in the academic record.

   https://dl.acm.org/doi/10.1145/3637528.3671900

4. **AirOps (2025)**

   Research analysing more than 21,000 brands. Source of the finding that 85% of brand mentions in AI search come from third-party content rather than the brand's own site.

   https://www.airops.com/report/the-influence-of-offsite-signals-in-ai-search

5. **Authoritas / PPA (2025)**

   Authoritas research, cited by the Professional Publishers Association via Sajeeda Merali, on AI Overviews reducing click-through rates by up to 57%.

   https://www.inpublishing.co.uk/articles/the-year-things-changed-again-26248

6. **BCG (2025)**

   Boston Consulting Group Black Friday Consumer Study, October 2025. Survey of 10,240 consumers across 10 countries on generative AI use during year-end sales events.

   https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap

7. **Benchmarking PDF Accessibility Evaluation (2025)**

   Benchmark dataset of scholarly PDFs with expert-validated accessibility annotations across 7 WCAG and PDF/UA criteria, with an LLM-based framework for accessibility assessment.

   https://arxiv.org/abs/2509.18965

8. **Cluttered to Clear (2025)**

   Research on improving web accessibility design for screen reader users in e-commerce using generative AI.

   https://arxiv.org/abs/2502.18701

9. **Deloitte (2026)**

   Deloitte 2026 Global Retail Industry Outlook. Source of the 9-in-10 retail leaders expectation and the 50% expecting collapse to a single AI-driven interaction by 2027.

   https://www.deloitte.com/mt/en/Industries/consumer/perspectives/global-retail-industry-outlook.html

10. **DMG Media (2025)**

DMG Media submission to the UK Competition and Markets Authority on AI Overviews and click-through rate impact. Source of the 89% desktop and 87% mobile decline figures on certain search terms.

https://www.searchenginejournal.com/impact-of-ai-overviews-how-publishers-need-to-adapt/556843/

11. **Duane Morris LLP (2026)**

Legal alert on the DOJ's Interim Final Rule extending ADA Title II digital accessibility deadlines by 1 year.

https://www.duanemorris.com/alerts/doj_extends_ada_title_ii_digital_accessibility_deadlines_one_year_0426.html

12. **Dubois, Dawson and Jaiswal (HBR June 2025)**

Harvard Business Review article: Forget What You Know About SEO. Here's How to Optimize Your Brand for LLMs. Source of the 58% (vs 25% in 2023) consumer migration figure and the 1,300% holiday season AI search referral surge.

https://hbr.org/2025/06/forget-what-you-know-about-seo-heres-how-to-optimize-your-brand-for-llms

13. **Enders Analysis / PPA (2025)**

Enders Analysis research, cited by the Professional Publishers Association, on UK media groups and search traffic decline across 2025.

https://pressgazette.co.uk/media-audience-and-business-data/uk-and-us-publishers-says-google-ai-is-harming-website-traffic/

14. **European Commission (2019)**

European Accessibility Act, Directive (EU) 2019/882. Application from 28 June 2025.

https://eur-lex.europa.eu/eli/dir/2019/882/oj

15. **Forrester (Lai et al, 2025)**

Forrester report: GenAI Forever Changes All Forms Of Search. Authored by Nikhil Lai with co-authors. Source of the analysis that the established measurement framework for search is shifting to AI-mediated metrics.

https://www.forrester.com/report/genai-forever-changes-all-forms-of-search/RES182189

16. **Government Digital Service (2018)**

GDS guidance: Why GOV.UK content should be published in HTML and not PDF.

https://gds.blog.gov.uk/2018/07/16/why-gov-uk-content-should-be-published-in-html-and-not-pdf/

17. **Google (2025)**

Google announcement on AI Overviews expansion to more than 200 countries and territories and more than 40 languages.

https://blog.google/products-and-platforms/products/search/ai-overview-expansion-may-2025-update/

18. **Google Search Central (2025)**

Google developer guidance on succeeding in AI search.

https://developers.google.com/search/blog/2025/05/succeeding-in-ai-search

19. **Hui, Lu and Zhang (NeurIPS 2024)**

Tsinghua University and National University of Singapore peer-reviewed paper introducing UDA: A Benchmark Suite for Retrieval Augmented Generation in Real-world Document Analysis. 2,965 documents and 29,590 expert-annotated Q&A pairs.

https://arxiv.org/abs/2406.15187

20. **Iliadis (JASIST 2023)**

Peer-reviewed journal article: One schema to rule them all: How Schema.org models the world of search. Journal of the Association for Information Science and Technology.

https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24744

21. **Information Commissioner's Office (2025)**

ICO guidance on the use of storage and access technologies, covering tracking pixels, link decoration, navigational tracking, device fingerprinting, web storage, and scripts and tags.

https://ico.org.uk/for-organisations/direct-marketing-and-privacy-and-electronic-communications/guidance-on-the-use-of-storage-and-access-technologies/

22. **Jackson Lewis P.C. (2026)**

Legal commentary on the DOJ Title II IFR and the HHS Section 504 deadline.

https://www.jacksonlewis.com/insights/doj-extends-public-entities-compliance-deadline-ada-related-website-accessibility-hhss-may-2026-deadline-still-looms

23. **Koreeda and Manning (2021)**

Hitachi America and Stanford University paper: Capturing Logical Structure of Visually Structured Documents with Multimodal Transition Parser. F1 score 0.951 vs 0.739 against a popular PDF-to-text tool.

https://arxiv.org/abs/2105.00150

24. **Lin (ChatDOC 2024)**

Paper: Revolutionizing Retrieval-Augmented Generation with Enhanced PDF Structure Recognition. Empirical evidence that PDF parsing accuracy materially affects retrieval-augmented generation quality.

https://arxiv.org/abs/2401.12599

25. **McKinsey (Silliman et al, October 2025)**

McKinsey Growth, Marketing & Sales Practice report: New front door to the internet: Winning in the age of AI search. Authored by Elizabeth Silliman, Julien Boudet, Kelsey Robinson, Desirae Oppong, and Nilay Shah. Source of the 16% brand-tracking figure, the 5 to 10% owned-content rule, and the projection of $750 billion in spending through AI search by 2028.

https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/new-front-door-to-the-internet-winning-in-the-age-of-ai-search

26. **NIST (2024)**

Artificial Intelligence Risk Management Framework: Generative AI Profile (NIST AI 600-1). Provenance, misinformation, and information integrity.

https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

27. **OpenAI (2025)**

OpenAI announcement on powering product discovery in ChatGPT and the role of structured merchant feeds.

https://openai.com/index/powering-product-discovery-in-chatgpt/

28. **Pettiette and Whitler (MIT Sloan, January 2026)**

MIT Sloan Management Review: Can Customers Find Your Brand? Marketing Strategies for AI-Driven Search. By Michael Pettiette (University of Houston) and Kimberly A. Whitler (University of Virginia Darden School of Business).
https://sloanreview.mit.edu/article/can-customers-find-your-brand-marketing-strategies-for-ai-driven-search/

29. **Puntoni (HBR February 2026)**
Harvard Business Review: AI Is Upending Marketing on Two Fronts. By Stefano Puntoni, Wharton School, University of Pennsylvania, and codirector of Wharton Human-AI Research.
https://hbr.org/2026/02/ai-is-upending-marketing-on-two-fronts

30. **Rahman and Finin (2017)**
Paper: Understanding the Logical and Semantic Structure of Large Documents. PDF rendering tools are not designed to support post-processing.
https://arxiv.org/abs/1709.00770

31. **Salesforce (2025)**
Salesforce 2025 holiday shopping data. Source of the $262 billion AI and agents share figure.
https://www.salesforce.com/uk/news/stories/2025-holiday-shopping-data/

32. **Section508.gov (2024)**
US federal guidance on PDFs and accessible content. Directs publishing toward HTML rather than PDF.
https://www.section508.gov/create/pdfs/

33. **Similarweb (Barenholtz, 2025)**
Similarweb research on zero-click searches, authored by Limor Barenholtz, Director of SEO and GEO at Similarweb. Zero-click search grew from 56% to 69% in the year following AI Overviews launch.
https://www.similarweb.com/blog/marketing/seo/zero-click-searches/

34. **UK Government (2010)**
Equality Act 2010.
https://www.legislation.gov.uk/ukpga/2010/15/contents

35. **UK Government (2018)**
Public Sector Bodies (Websites and Mobile Applications) (No. 2) Accessibility Regulations 2018.
https://www.legislation.gov.uk/uksi/2018/952/contents/made

36. **US Department of Justice (2024)**
Title II of the Americans with Disabilities Act, Final Rule on Web Content and Mobile Application Accessibility, 24 April 2024.
https://www.ada.gov/resources/web-rule-first-steps/

37. **US Department of Justice (2026)**
Interim Final Rule, 91 Fed. Reg. 20902, 20 April 2026. Extension of Compliance Dates for Nondiscrimination on the Basis of Disability; Accessibility of Web Information and Services of State and Local Government Entities.
https://www.federalregister.gov/documents/2026/04/20/2026-07663

38. **Volpini et al. (WordLift 2026)**

Structured Linked Data as a Memory Layer for Agent-Orchestrated Retrieval. 29.6% accuracy improvement in standard RAG, 29.8% in agentic RAG, across 2,443 query evaluations in 4 industry verticals.
https://arxiv.org/abs/2603.10700

39. **Walsh, Boudreaux et al. (2023)**

Maximizing Equitable Reach and Accessibility of ETDs. Direct study using AI to convert PDF documents to machine-readable HTML with semantic tags, with the explicit dual aim of accessibility and machine-readability.
https://arxiv.org/abs/2310.18427

Paper brought to you by accessapdf;

Your problem

Most organisations carry large numbers of PDFs accumulated over years of web publishing. They sit across websites and legacy structures, with no clear picture of what exists, where it lives, or who owns it. The first task is finding them. The second is making them work for the people, search engines, and AI systems now reading them.

The reach of that content has changed. AI search and AI agents now read, summarise, and present published material on a visitor's behalf. Where content is inaccessible, outdated, duplicated, or unclear, the picture they form of the organisation reflects those gaps. Manual approaches cannot keep pace with the volume.

What the platform does

AccessaPDF converts the PDFs linked across a website into inclusive web content that people, search engines, and AI systems can read.

The work begins with discovery, scanning the site and generating a structured catalogue of the document estate. For the first time, organisations see what they hold, where it sits, and the scale of the work involved. Linked PDFs are converted into inclusive HTML, served alongside the original through the secure tag.

Ongoing, organisations define what is correct for their own content. Approved positioning, terminology, and facts are held centrally, and the platform flags inaccuracies in material already online. The original PDF is never altered, and the process is consistent, documented, and repeatable.

Why it matters

Published content now has more readers than people alone. Search engines, AI search, and AI agents read the same material, summarise it, and represent the organisation on a visitor's behalf.

PDFs sit at the centre of that exposure. Left in place, they shape how people, machines, and decisions outside the organisation see what is published. A structured, ongoing process keeps inclusive content available as documents change, new ones appear, and the audience reading them grows.

That is what AccessaPDF is built to support.

Visit

https://accessapdf.aaanow.ai/