

THE STANDARD, IN ONE PAGE-AND-A-BIT

# How should a page be built so machines **quote it correctly?**

Lead with the answer, structure the page so extraction cannot mangle it, and say the same thing everywhere. In practice that is eight rules: a question-shaped h1, an answer-first opening, semantic headings with real text, a visible FAQ, valid JSON-LD on stable URLs, one clear CTA, an llms.txt index at the root, and entity consistency. This is the standard TBSCG applies to every page it ships, published for anyone to adopt.

*Nothing here needs new tooling or a replatform. Each rule stands alone.*

01 · WHY SCRAPE BIAS

## Extraction has a bias. Build for it.

Answer engines do not read a page the way a person does. They crawl it, extract candidate passages, and assemble an answer from the passages that survive. That process favours pages that answer first, structure cleanly and mean exactly one thing. A page that buries its point under three screens of brand prose may still rank in a list of links; it rarely survives extraction intact.

Ambiguity is a tax. Every place a page makes the machine guess, two names for the same service, a claim that only renders after JavaScript runs, an FAQ hidden inside an accordion the crawler never opens, is a place the answer can come out wrong, or come out citing someone else. The standard below removes the guessing.

02 · THE STANDARD · EIGHT RULES

## Eight rules, adoptable today.

01

### One h1, phrased as the buyer's question

Every page answers exactly one question, and the h1 is that question in the buyer's words. Machines treat the h1 as the strongest signal of what a page is for. One page, one question, one h1.

03

### Semantic headings and real text

Structure lives in h2 and h3 elements, not styled divs, and the content is real text in the markup. Nothing load-bearing sits behind JavaScript: if a claim only exists after a script runs, assume the machine never reads it.

02

### An answer-first opening

The first paragraph under the h1 answers the question fully, on its own, before any context or positioning. Assume the extractor takes that paragraph and nothing else. If it survives alone, the page survives extraction.

04

### A visible, non-collapsed FAQ

Five questions per page, each an h3 with its answer in plain text below it. No accordions, no tabs, no click-to-reveal. The questions mirror what buyers actually ask, because those are the prompts the engines receive.

## 05

**Valid JSON-LD on stable URLs**

Service or product schema plus FAQPage, validated, on production URLs with shallow paths that do not change. Structured data hands the machine facts instead of prose to interpret; URL churn throws those facts away.

## 07

**An llms.txt index at the domain root**

A plain-text file that states what the organisation is in one line and lists the canonical pages, so a crawler on a token budget reads the curated version first. Ours is live at [www.tbscg.com/llms.txt](http://www.tbscg.com/llms.txt).

## 06

**One clear CTA**

Each page ends with a single next step. Competing calls to action read as noise to a person and as ambiguity to a machine summarising what the page wants the reader to do.

## 08

**Entity consistency**

The same firm name, the same service names, the same one-line descriptions, everywhere: on the page, in the schema, in llms.txt, in third-party profiles. Machines resolve entities by repetition; every variant spelling splits your identity.

**HOW TO KNOW A PAGE MEETS THE STANDARD**

- 01** Read the first paragraph in isolation and ask whether it answers the h1 on its own.
- 02** Validate the JSON-LD.
- 03** View the page with JavaScript disabled and check that nothing load-bearing disappears.
- 04** Then ask an answer engine the h1's question and see what comes back.

**ADOPTING IT**

You do not have to adopt all eight rules at once. Each rule stands alone and can be applied page by page. Most teams start with the question-shaped h1 and the answer-first opening on their highest-value pages, then add the FAQ and the schema, then publish llms.txt once the canonical pages are settled. The standard is compatible with conventional SEO and usually improves it, because search engines also reward clean structure and structured data.

Published by TBSCG and applied to every page it ships. If you want to know how far your estate currently sits from the standard, the DXP Value Survey scores machine visibility inside its AI-readiness dimension, in four weeks at a fixed price. Start at [www.tbscg.com/practice/survey.html](http://www.tbscg.com/practice/survey.html).

*Standards work better in the open. Take what is useful.*