Brandella Journal

Crawlability Checks checklist

Project: _____ Date: _____

- [ ] robots.txt returns 200, is under 500 KiB and uses only user-agent, allow, disallow and sitemap
- [ ] No page you want indexed is blocked by robots.txt, for Googlebot or for the AI search bots you allow
- [ ] Any named bot group repeats every * rule that bot still needs
- [ ] CSS, JavaScript and fonts needed to render pages are not blocked
- [ ] No noindex in meta robots or X-Robots-Tag on pages you want indexed
- [ ] Pages with noindex are not blocked, so Google can see the noindex
- [ ] Every page you want indexed returns 200 and has a self referencing canonical
- [ ] Canonicals point to live, indexable URLs, never to redirects or 404s
- [ ] Sitemaps list only 200, indexable, canonical URLs, each file under 50,000 URLs and 50 MB
- [ ] Sitemaps use absolute URLs, are listed in robots.txt or submitted, and lastmod is accurate
- [ ] Every page you want indexed has internal links and sits within 3 clicks of the home page
- [ ] Main content is in the raw HTML, not only after JavaScript runs
- [ ] Filter, search, tag, staging and old URLs are handled with noindex, canonicals, 404s or passwords on purpose
- [ ] Your policy for every search and AI bot matches what robots.txt and your CDN really do
- [ ] The Page indexing and Sitemaps reports are checked weekly and robots.txt changes are compared

Full kit and guide: https://brandella.in/resources/seo-aeo-assets/crawlability-checks/