

# Robots.txt Checklist

- ■ Ensure the robots.txt file is placed in the root directory of the website (example.com/robots.txt).
- ■ Confirm the file is publicly accessible (test via browser).
- ■ Specify User-agent rules (e.g., Googlebot, Bingbot, or \* for all).
- ■ Block sensitive or duplicate content from crawling (e.g., /admin/, /cart/, /tmp/).
- ■ Avoid blocking essential resources (e.g., CSS, JS, images needed for rendering).
- ■ Include sitemap location at the end of the file (e.g., Sitemap: <https://example.com/sitemap.xml>).
- ■ Do not block important SEO pages (homepage, product pages, blog posts).
- ■ Use Disallow carefully—avoid blocking entire site by mistake (/).
- ■ Test rules in Google Search Console's robots.txt tester.
- ■ Keep file size under 500KB (Googlebot limit).
- ■ Ensure no typos or invalid syntax (case-sensitive).
- ■ Update file whenever site structure changes.
- ■ Add Crawl-delay only if server performance requires it.
- ■ Allow important bots (e.g., AdsBot-Google, Mediapartners-Google for ads).
- ■ Check for multiple robots.txt files (only one per domain is valid).
- ■ Verify HTTPS and HTTP versions both serve the same robots.txt.