

candour

# Robots vs Noindex usage



Trainer:

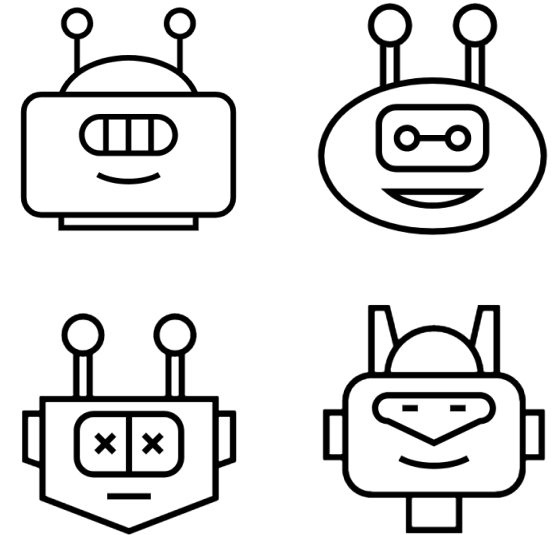
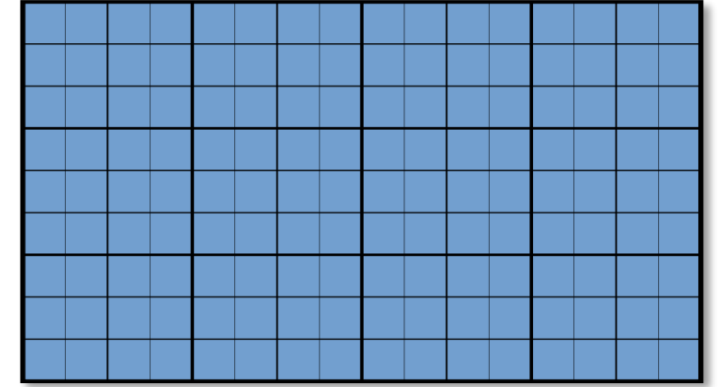
Mark Williams-Cook  
Candour



# Robots and Noindex usage

Robots.txt **stops crawling**

Noindex tag **prevents indexing**



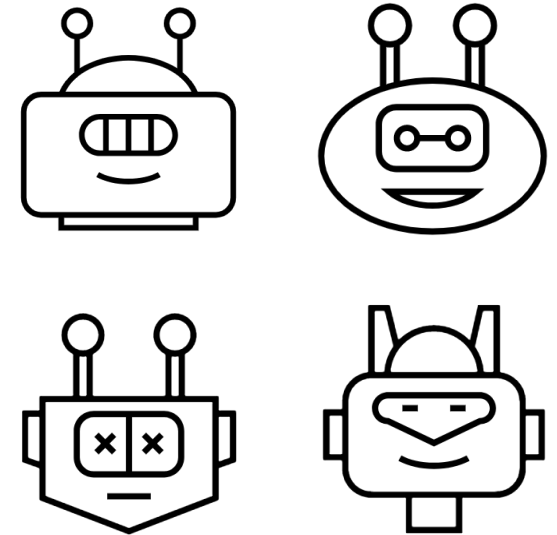
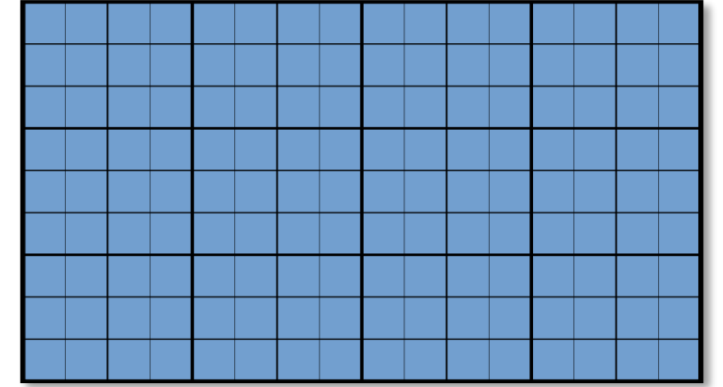
# Robots and Noindex usage

Robots.txt **stops crawling**

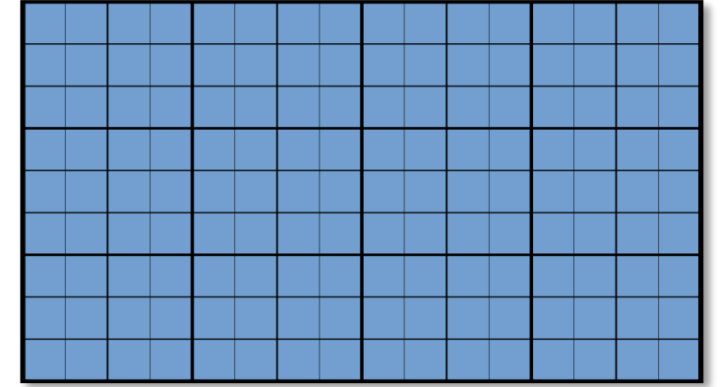
To prevent bots wasting resources on pages that are not important

Noindex tag **prevents indexing**

When we don't want pages appearing in search results



# Key concept



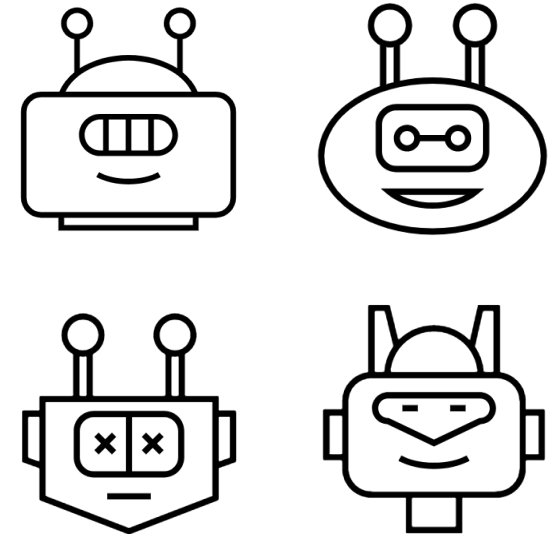
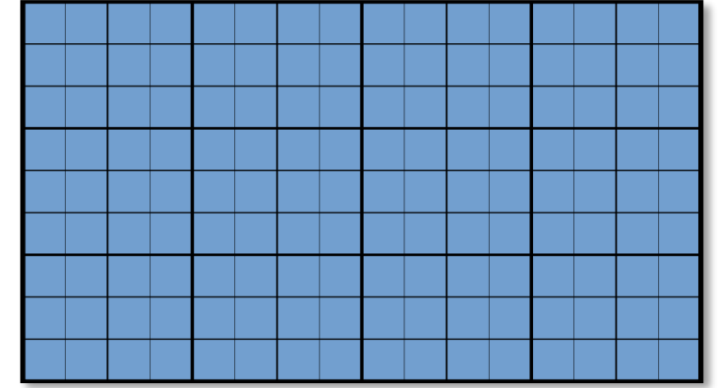
While the robots.txt will not stop indexing, it will make it far less likely to occur.



# Example: Staging site

Our live website: [withcandour.co.uk](https://withcandour.co.uk)

Our staging website: [staging.withcandour.co.uk](https://staging.withcandour.co.uk)



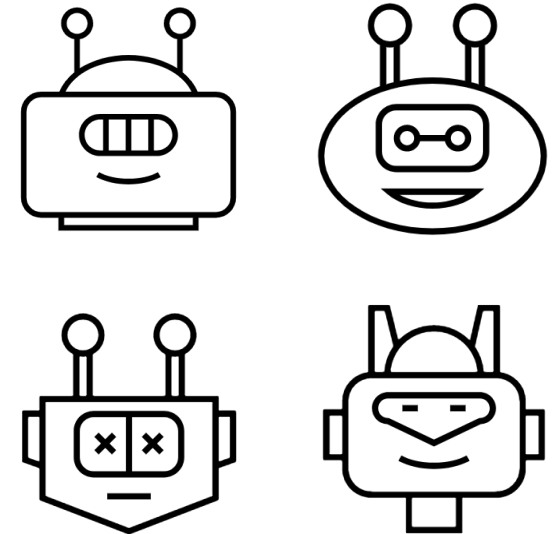
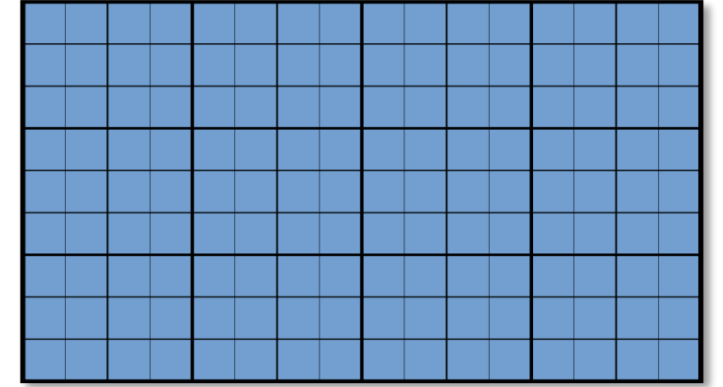
# Challenge



In the case of development or staging websites, should we be adding them to robots.txt or should we be using the noindex tag?

# Example: Staging site

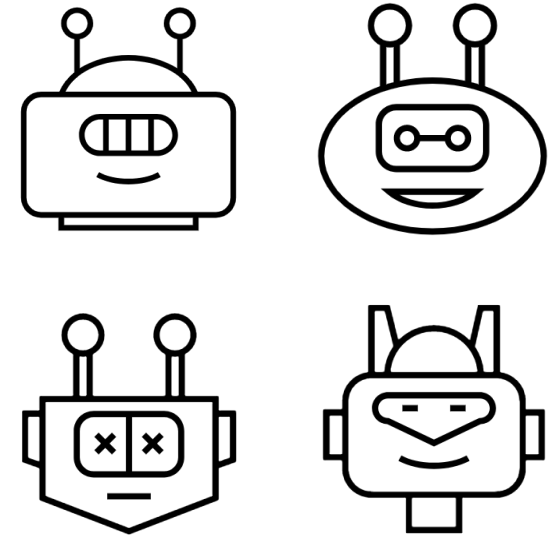
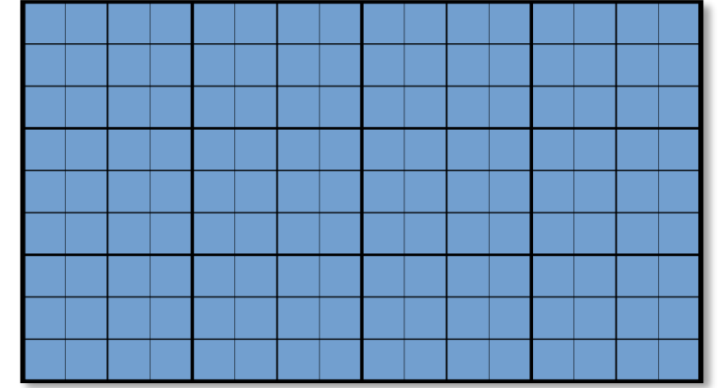
In the example of a staging site we would use a **noindex tag** because:



# Example: Staging site

In the example of a staging site we would use a **noindex tag** because:

- 1) We definitely don't want this appearing in search

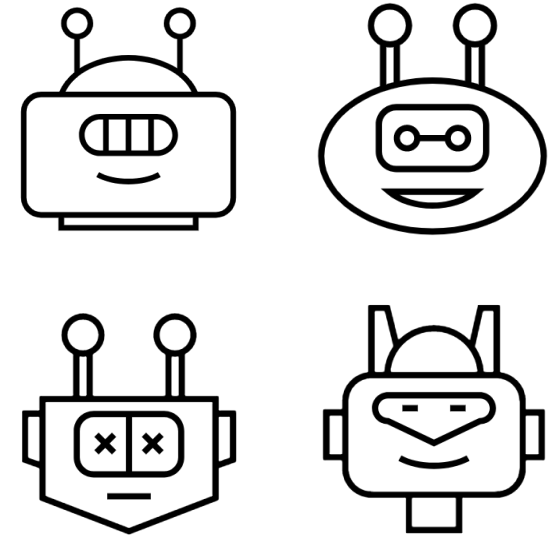
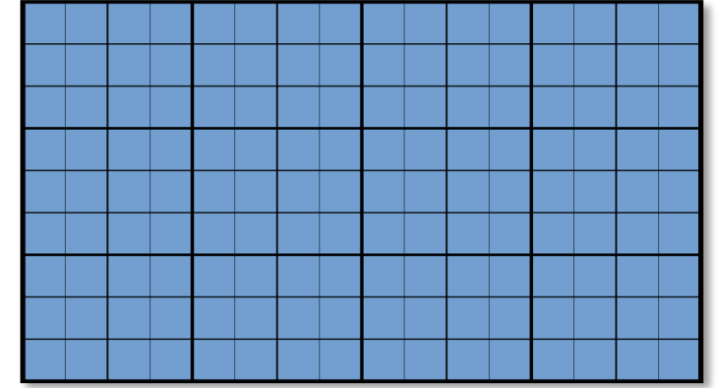




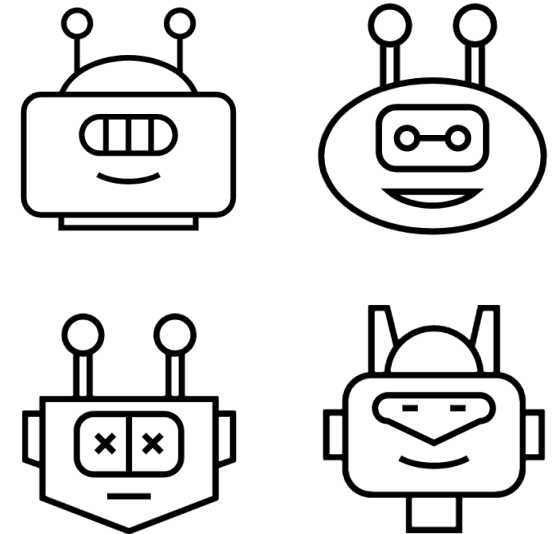
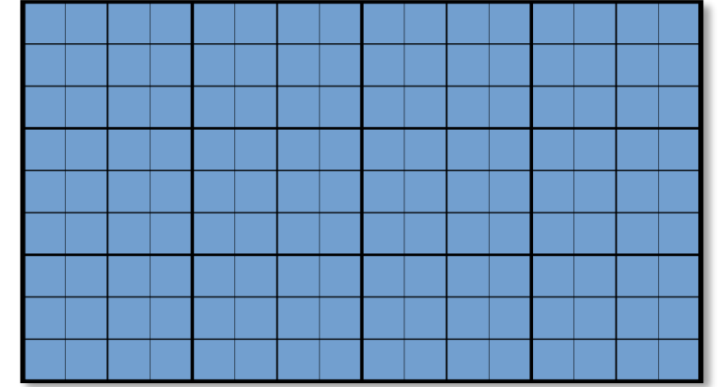
# Example: Staging site

In the example of a staging site we would use a **noindex tag** because:

- 1) We definitely don't want this appearing in search
- 2) There are unlikely to be any links to this site for us to stop crawling



# Spider traps

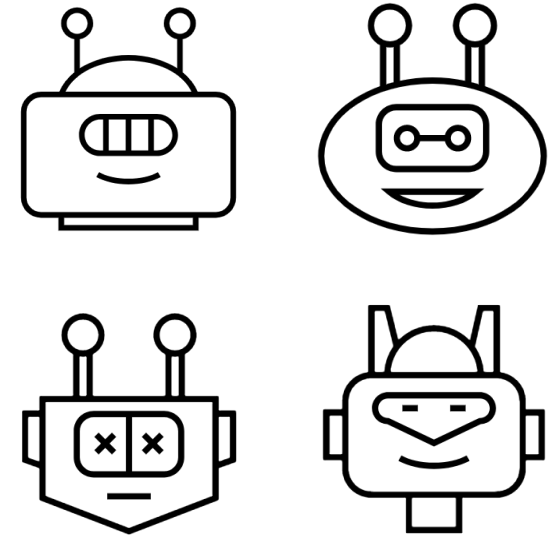
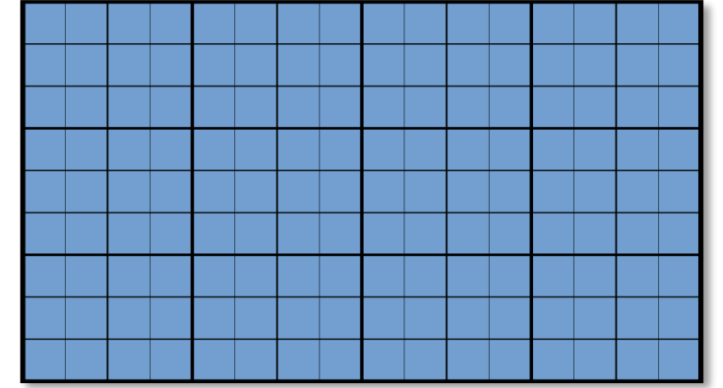


---

Robots and Noindex usage

# Spider traps

When spiders (robots) get “trapped” in a large or never-ending set of pages, usually of low-value.

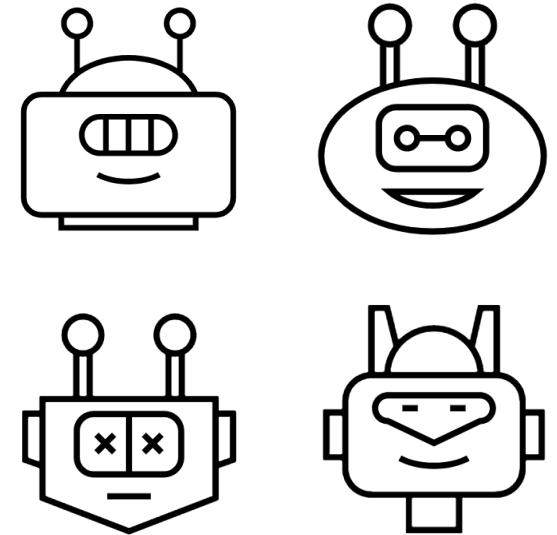
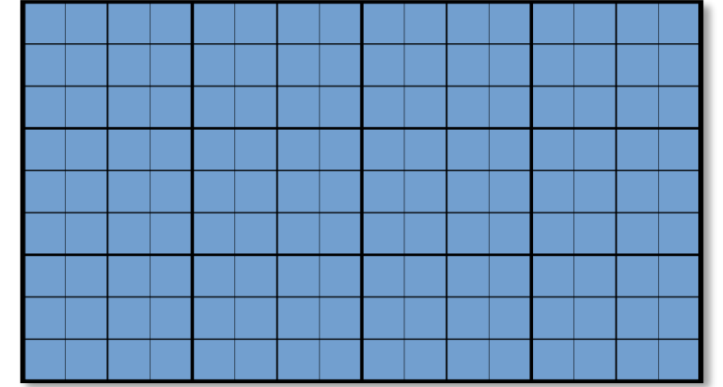


# Challenge



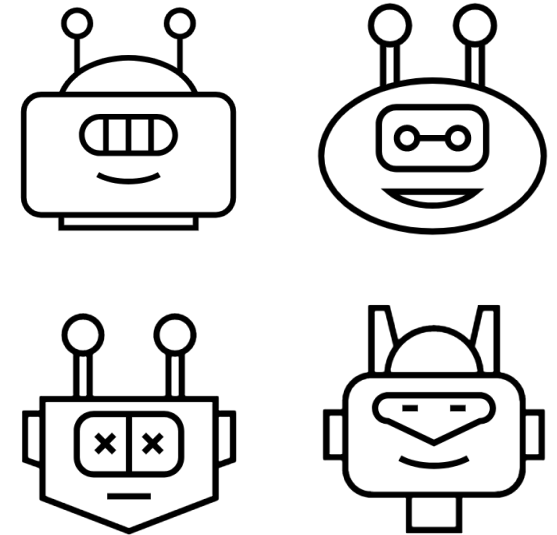
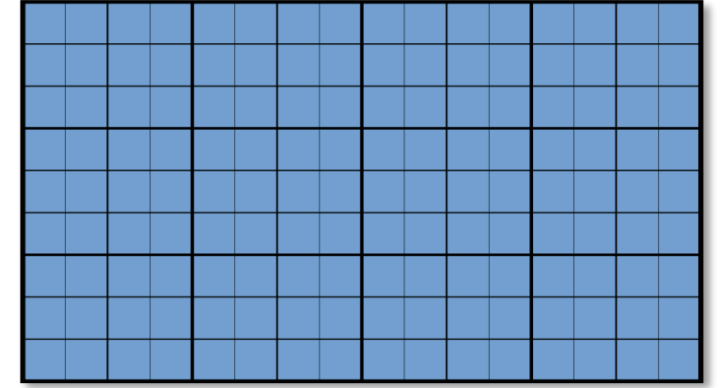
How could we tell if we have a ‘spider trap’ on our site?

# Checking for spider traps



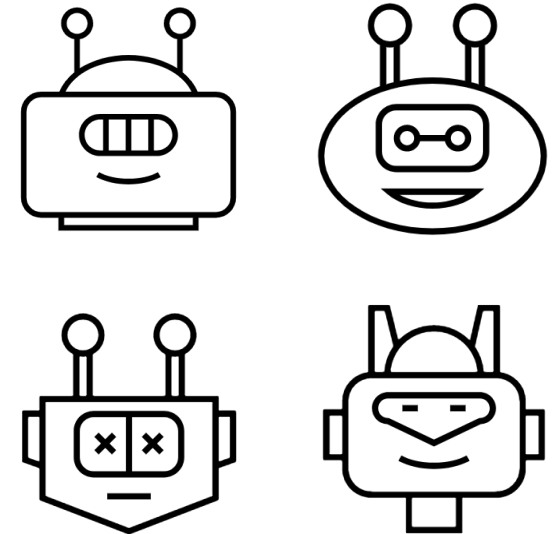
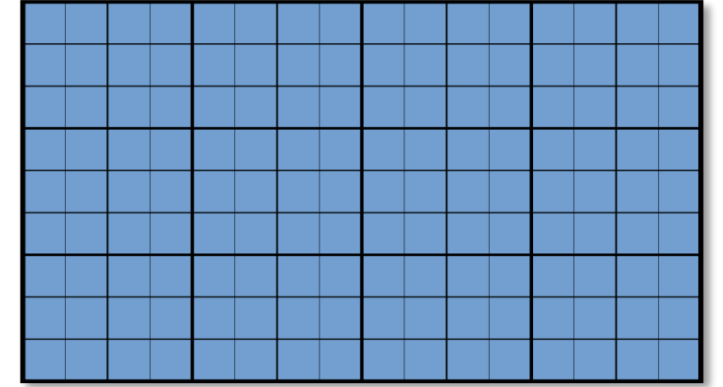
# Checking for spider traps

1) Use the site: operator in Google



# Checking for spider traps

- 1) Use the **site:** operator in Google
- 2) Crawl the site with Screaming Frog



# Challenge



Run a partial Screaming Frog crawl of the Fig Leaves site, ignoring the robots.txt, then compare the robots.txt file and look at the rules they have set up to fix this issue.