Machine Learning engineer

The goal of this task is to understand your approach to tackling challenges that occur in our day-to-day work. The challenge at hand is something that might get on your desk if you start working at PhishAR.

Challenge

Implement a simple reverse image search for screenshots containing full-screen browser contents(*). Your solution should take an image as input and return a fully qualified domain name (FQDN, e.g. "cnn.com") of the website displayed on the browser.

Your solution should work on these websites:

- https://www.theguardian.com/
- https://www.spiegel.de/
- https://cnn.com
- https://www.bbc.com/
- https://www.amazon.com/
- https://www.ebay.com/
- https://www.njuskalo.hr/
- https://www.google.com/
- https://github.com/
- https://www.youtube.com/

Be careful of images belonging to none of these websites!

Bonus points for:

⭐ your solution being packaged in a docker container

⭐ deploying a simple web application that takes an image as input and displays the detected website's FQDN

⭐⭐ your solution working for top 1000 Alexa-ranked websites

⭐⭐⭐ your solution working on photographs of screens

(*) see last page for an example

phishAR

info@phishar.com
9400 Garsington Road,
Oxford, OX4 2HN, United Kingdom

Deliverables

1. Your source code (git), be sure to include a short explanation of your approach in the `README.md` file, explain how you got your dataset
2. A sample of your dataset
3. (optional) A link to the deployed web application anywhere online (e.g. Heroku)
4. (optional) Anything additional you would like to show us/further explain

What we'll be looking at

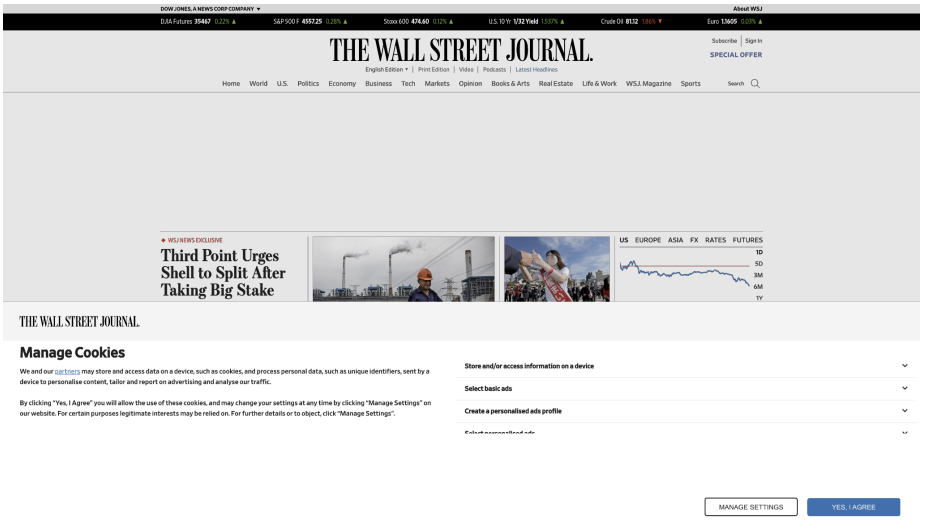
1. Your approach to the problem at hand
2. Your setup of the application and/or architecture
3. Your preference of technologies
4. The usability of your solution
5. Your code hygiene

Additional information

1. This is a very open-ended problem and we don't expect your solution to work perfectly. Deep-learning approaches are welcome but not expected.
2. If you have any additional questions, please let us know
3. The code you submit will be used only for assessing your application

info@phishar.com
9400 Garsington Road,
Oxford, OX4 2HN, United Kingdom

Example input:



Expected output:

wsj.com