

Codebook

(refer to sec 6 in our WWW paper)

Phish:

1. Layout: (appearance, location)
 - Bare webpage
 - Header
 - Footer
 - Design
 - Button
 - Text boxes
2. Functionality: (webpage doesn't make sense given the context of the service)
 - Third-party service (e.g., many email options)
 - Captcha
 - Link (e.g., redirect link)
 - Authentication
 - i. Auto Login
 - ii. Email (e.g., pre-filled)
 - iii. Password visible
 - iv. 2FA
 - v. Ask for sensitive information (e.g., credit card)
 - vi. Account verification (e.g., do not have email verifying)
3. Content:
 - Text
 - i. Typo
 - ii. Grammar
 - iii. Text style
 - iv. Capitalization
 - v. Language (e.g., the webpage includes other languages)
 - vi. Footer
 - Lack of info: (e.g., no contact info)
 - Advertisement
 - Privacy policy
 - Copyright/trademark
 - Sponsor
 - Image(quality of the image)
 - i. Background
 - ii. Footer image
 - Symbol
 - i. Lock
 - ii. Security certificate
4. Outdated

5. URL (e.g., ask for URL)
6. Too good to be true (or Easy to replicate an existing website)
7. Brand:
 - Unfamiliar (e.g., do not know this brand)
 - reputation
 - Confuse brand (e.g., Multi-brand)
8. Logo (e.g., logo distortion)

Unsure

Legitimate:

1. Layout: (appearance, location)
 - Overall consistent
 - Design
 - Footer
2. Functionality:
 - Cookie
 - Captcha
 - Link
 - Authentication
 - i. Email (pre-filled)
 - ii. Password verification
 - iii. No sensitive info
 - iv. Multiple login options
3. Content:
 - Text
 - i. Spell: (e.g., correct spelling)
 - ii. Grammar
 - iii. Text style
 - iv. Footer
 - Copyright
 - Advertisement
 - Insurance
 - Weather
 - Lots of info
 - Claim
 - Image
 - Symbol
 - i. Lock
 - ii. Security certificate
6. URL
7. Brand:
 - familiar (e.g., know this brand)
 - reputation
9. Logo
8. Simplicity (e.g., website is too simple to be phish)