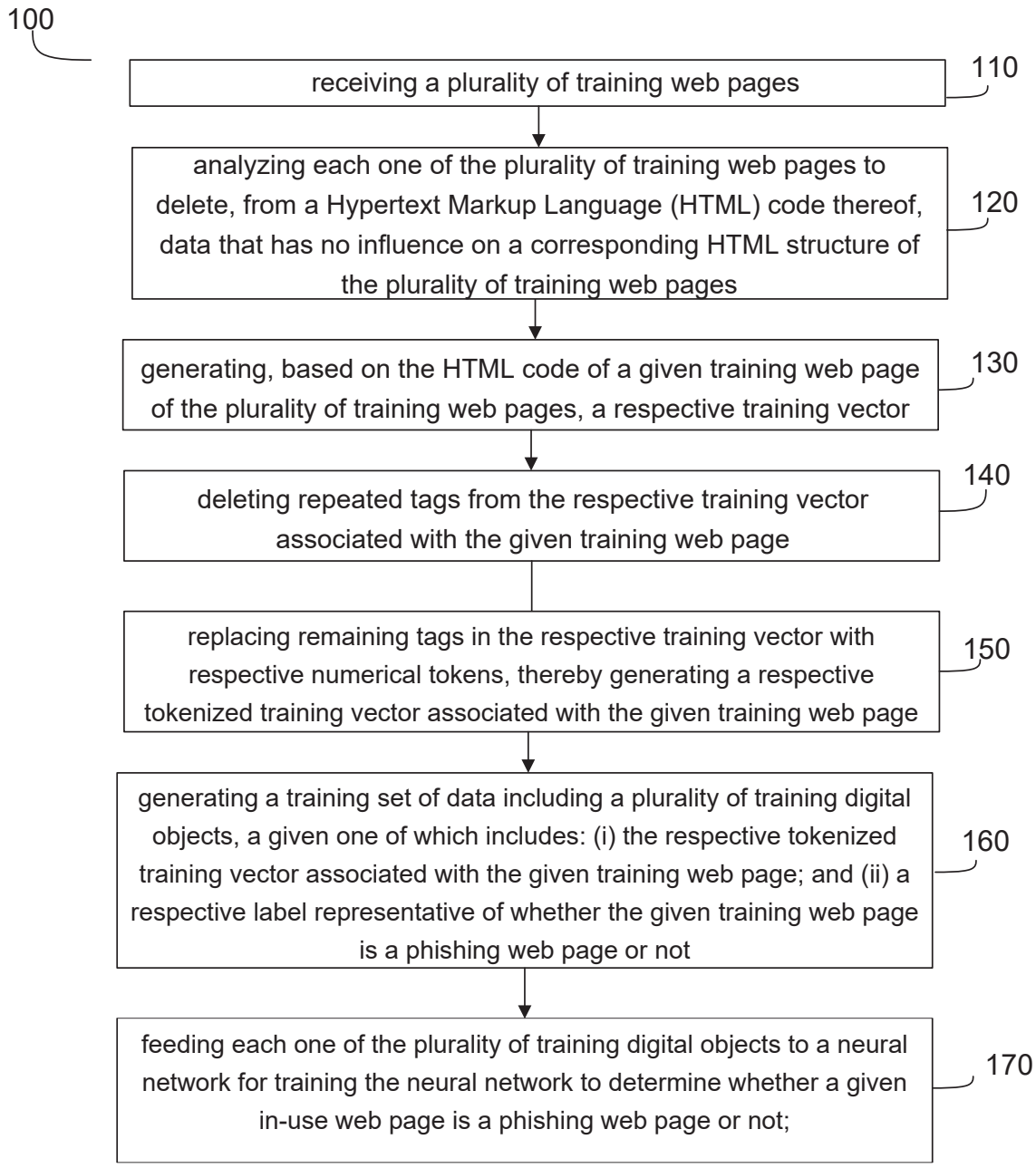

100
receiving a plurality of training web pages
110
analyzing each one of the plurality of training web pages to delete, from a Hypertext Markup Language (HTML) code thereof, data that has no influence on a corresponding HTML structure of the plurality of training web pages
120
generating, based on the HTML code of a given training web page of the plurality of training web pages, a respective training vector
130
deleting repeated tags from the respective training vector associated with the given training web page
140
replacing remaining tags in the respective training vector with respective numerical tokens, thereby generating a respective tokenized training vector associated with the given training web page
150
generating a training set of data including a plurality of training digital objects, a given one of which includes: (i) the respective tokenized training vector associated with the given training web page; and (ii) a respective label representative of whether the given training web page is a phishing web page or not
160
feeding each one of the plurality of training digital objects to a neural network for training the neural network to determine whether a given in-use web page is a phishing web page or not;
170

FIG.1

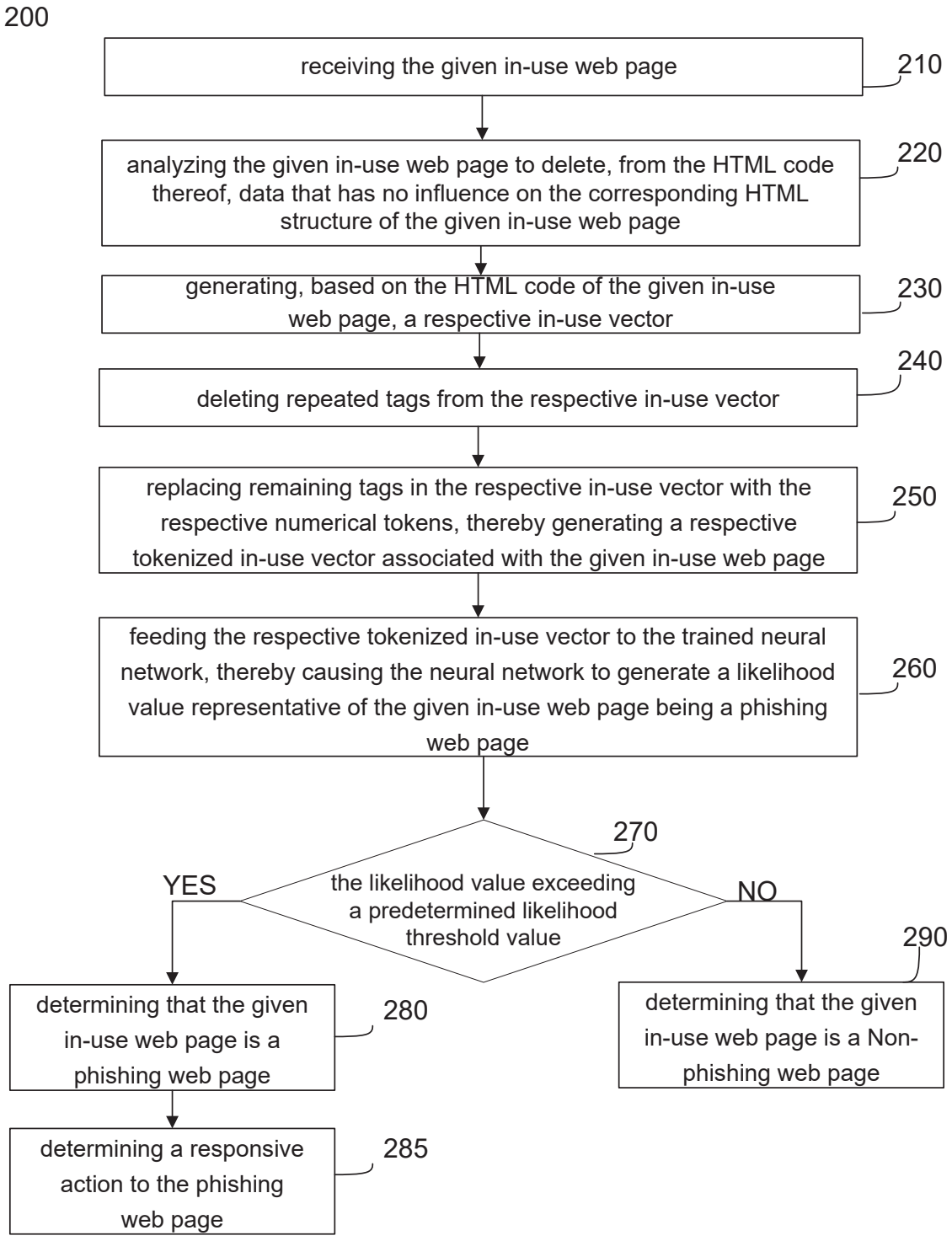

200
receiving the given in-use web page
210
analyzing the given in-use web page to delete, from the HTML code thereof, data that has no influence on the corresponding HTML structure of the given in-use web page
220
generating, based on the HTML code of the given in-use web page, a respective in-use vector
230
deleting repeated tags from the respective in-use vector
240
replacing remaining tags in the respective in-use vector with the respective numerical tokens, thereby generating a respective tokenized in-use vector associated with the given in-use web page
250
feeding the respective tokenized in-use vector to the trained neural network, thereby causing the neural network to generate a likelihood value representative of the given in-use web page being a phishing web page
260
270
the likelihood value exceeding a predetermined likelihood threshold value
YES
NO
determining that the given in-use web page is a phishing web page
280
determining that the given in-use web page is a Non-phishing web page
290
determining a responsive action to the phishing web page
285

FIG. 2




```
<!doctype html>
<html>
<head>
    <title>Example Domain</title>
    <meta charset="utf-8" />
    <meta http-equiv="Content-type" content="text/html;
charset=utf-8" />
    <meta name="viewport" content="width=device-width, initial-
scale=1" />
    <style type="text/css">
    body {
        background-color: #f0f0f2;
        margin: 0;
        padding: 0;
        font-family: -apple-system, system-ui, BlinkMacSystemFont,
"Segoe UI", "Open Sans", "Helvetica Neue", Helvetica, Arial, sans-
serif;
    } div
    {
        width: 600px;
        margin: 5em auto;
        padding: 2em;
        background-color: #fdfdff;
        border-radius: 0.5em;
        box-shadow: 2px 3px 7px 2px rgba(0,0,0,0.02);
    }
    a:link, a:visited {
        color: #38488f;
        text-decoration: none;
    }
    @media (max-width: 700px) {
        div {
            margin: 0 auto;
            width: auto;
        }
    }
    </style>
</head>

<body>
<div>
    <h1>Example Domain</h1>
    <p>This domain is for use in illustrative examples in documents.
You may use this domain in literature without prior coordination or asking for
permission.</p>
    <p><a href="https://www.iana.org/domains/example">More
information...</a></p>

</div>
</body>
</html>
```

FIG. 3A



FIG. 3B

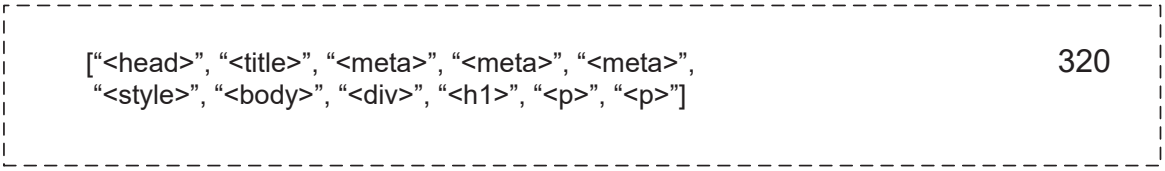


FIG. 3C

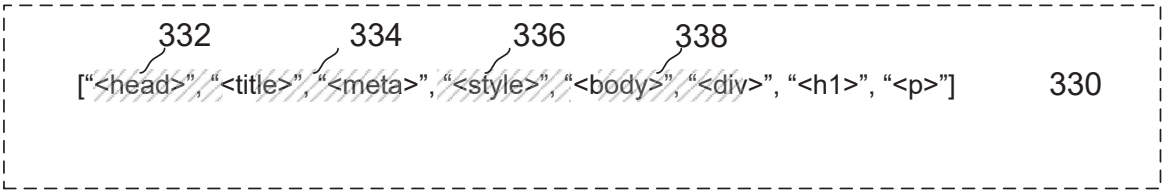


FIG. 3D



FIG. 3E

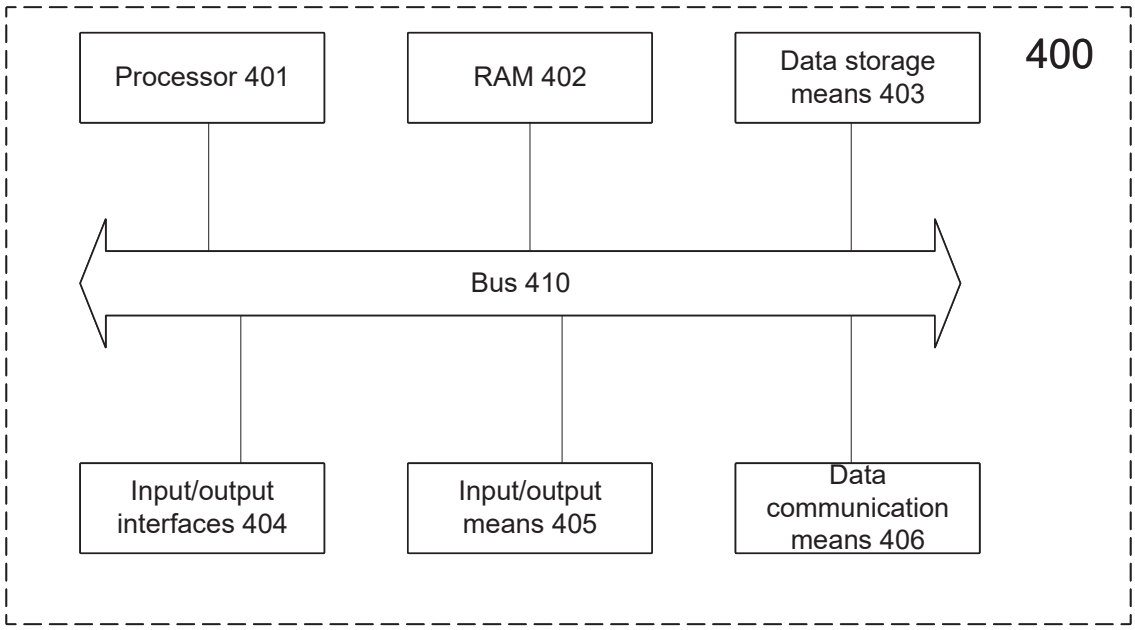

400
Processor 401
RAM 402
Data storage means 403
Bus 410
Input/output interfaces 404
Input/output means 405
Data communication means 406

FIG. 4