Webの自動巡回とDNS情報を利用した大規模言語モデルを用いたフィッシングサイトの検出

竹下 駆[1,a)]　黄 緒平[2,1,b)]　伊藤 彰則[2,c)]

概要：フィッシングサイトの報告件数も年々上昇し，手口も巧妙化しており判別がより難しくなっている．複雑な推論を行うことのできる大規模言語モデルを用いてフィッシングサイトを検出する手法に関する研究はまだ少ない．本研究では Web サイトの自動巡回により HTML，URL，スクリーンショットから OCR を抽出したテキストに加えドメイン名やサーバーの情報について収集した情報に基づいたプロンプトを使用し，大規模言語モデル（LLM）を用いてフィッシングサイト検出の精度向上を目指す．本研究では PhishTank から用意したフィッシングサイトのデータセット約 800 件と約 300 件の有名企業 URL を用いて実験を行った．得られた結果は本手法の適用可能性と改良の余地を示しており，今後の発展に向けた貢献となる．

キーワード：大規模言語モデル，フィッシングサイト，自動巡回，DNS

Detection of Phishing Websites Using Large Language Models with Automated Web Crawling and DNS Information

KAKERU TAKESHITA[1,a)]　XUPING HUANG[2,1,b)]　AKINORI ITO[2,c)]

Abstract: The number of reports of phishing sites is increasing every year, and the methods used are becoming more sophisticated, making it more difficult to distinguish them. There is still little research on methods for detecting phishing sites using large-scale language models capable of complex reasoning. In this study, we aim to improve the accuracy of phishing site detection by using prompts based on information collected through automatic web crawling, including HTML, URLs, text extracted from screenshots using OCR, domain names, and server information, and by utilizing large-scale language models (LLMs). In this study, experiments were conducted using approximately 800 phishing site datasets provided by PhishTank and approximately 300 URLs from well-known companies. The results demonstrate the applicability of the proposed method and indicate room for improvement, thereby contributing to future developments.

Keywords: Large-scale language models，Phishing sites，automatic calling，DNS

1. はじめに

フィッシングサイトは利用者を欺き，ID，クレジットカード情報，パスワード，住所，氏名等の個人情報を抜き出すサイトである．フィッシングサイトは正規のサイトを模倣してなりすまし，利用者に正規のサイトであると思い込ませたり，偽のマルウェア感染警告や架空の請求を装うなど主に人の恐怖や緊急性を煽ることでを活用することや，偽の報酬を提示することででユーザーの関心を引き情報を入力させる [1][2][3]．近年のサイバーセキュリティの分野では大規模言語モデル（LLM）を用いたアプローチが注目されている [4]．LLM は様々な情報を組み合わせ複合的に

[1] 島根大学総合理工学部
Interdisciplinary Faculty of Science and Engineering, Shimane University
[2] 東北大学工学研究科
Graduate School of Engineering, Tohoku University
[a)] s225033@matsu.shimene-u.ac.jp
[b)] huang@cis.shimane-u.ac.jp
[c)] aito.spcom@tohoku.ac.jp

推論を行うことができることからこれまでとは違う手法でのフィッシングサイトの検出の可能性が期待されている．

1.1 関連研究

従来の LLM を用いたフィッシングサイトの検出ではフィッシングサイトを特徴づける要素である（1）公式のロゴやブランディングを使用して正規のサービスや企業を模倣すること，（2）SE 技術を利用してユーザの行動を心理的に操作することである．主にユーザーに表示される画面に関する情報からフィッシングサイトを検出しようとする手法が主である [5]．小出らの研究によると URL，スクリーンショット，HTML の情報を入力することによって LLM を用いてフィッシングサイトを検出する実験を行い GPT-4V を用いて精度 98.7%を達成した．また吉田らの研究では Whois 情報，SSL 証明書情報，DNS サーバーへの問い合わせにより DNS グラフを作成し LLM に渡すことでフィッシングサイトの判別を行う手法を提案している [6]．以前の研究では複合的な情報 [7] を LLM に渡すことで判別の精度の向上を図って行っていたが，本研究ではいくつもの情報を渡すのではなく，一つずつ渡すことでどのような情報が判別精度の向上に役立っているかについて調べる．

2. 提案手法

フィッシングサイト検出のために使用したプロンプトは先行研究 [8] で提案されている ChatPhishDetector を参考にして作成する．ChatPhishDetector は WEB クローラーを使用して URL にアクセスし，到達した Web サイトからスクリーンショット画像，HTML，URL を取得する．スクリーンショットした画像から光学文字認識 (OCR) を行いテキストを抽出する．取得した情報に基づいて LLM への入力プロンプトを作成し，その Web サイトがフィッシングサイトかどうかを判断するシステムである．本研究では上記の情報に加えて Whois 情報を取得しそれを基にプロンプトを改良して LLM に判定させる場合，nslookup による DNS 情報を取得しそれを基にプロンプトに書き加え LLM に判定させる場合，python のライブラリ ssl を用いて取得した SSL 証明書の情報を書き加え LLM に判定させる場合，Whois，nslookup，SSL のすべてを利用した場合，tor ブラウザを利用した場合に変化はあるのかを確認するため，nslookup+Tor ブラウザのの 5 パターンで検証を行う．

2.1 プロンプト

今回利用したプロンプトテンプレートは**図 1** のとおりであり，先行研究 [8] で使用されたプロンプトテンプレートに一部加筆修正して使用している．フィッシングサイト検出タスクは，特定の推論プロセスの実行を促進する 4 つのサブタスクに分割される．これらのサブタスクは以下の通りである：1. ユーザを欺く SE 技術が含まれているかを分析する．フィッシングサイトに典型的な SE 技術（懸賞当選，マルウェア感染警告，アカウントや配達の問題）を例示する．2. Web サイトのブランド名を抽出する．フィッシングサイトは正規サイトから HTML や画像などのリソースを複製して作成されることがあるため，HTML のみに基づいて真正性を判断するのではなくドメイン名が正規のものと一致するかを検証する．3. Web サイトがフィッシングサイトであるかどうかを判断し，その根拠を明示的に記述する．詳細に説明の生成により，応答の正確性を向上させ，事後の解析者による確認を容易にすることを意図している．4. JSON 形式の出力を生成する．phishing はフィッシング判定結果，brands が Web サイトを代表するブランド，suspicious domain がドメイン名の不審さ，phishing score が 0 から 10 の範囲のスコアを表す．このプロンプトに抽出した OCR，スクリーンショット画像，URL を代入することでプロンプトテンプレートを完成させる．Whois 情報，nslookup 情報，ssl 情報に関するプロンプトは**図 2**Whois 情報，組織名，作成日，およびネームサーバーに不審な点がないか確認してください．**図 3**DNS，組織名，作成日，およびネームサーバーに不審な点がないか確認してください．**図 4** 取得した証明書情報を確認し，不審な点がないか確認してください．また，証明書の記載内容が予測されたブランドの情報と一致しているか確認してください．
上記の文章をプロンプトのサブタスクにそれぞれ書き加える．

3. 評価実験

先行研究ではすべての情報を使い精度の向上を図っていたが, 本研究で Whois 情報, DNS 情報, SSL 情報を一つずつ利用することで，LLM を用いてフィッシングサイトの検出を行うときにどの情報がもっとも大きく影響しているのかを調べる．本研究では LLM は gemini-2.0-flash-lite-preview を使用して実験を行う．

3.1 データセット

今回使用したデータセットは PhishTank から用意したデータセットである．PhishTank は利用者がフィッシングサイトの URL を報告し, 報告した利用者とは別の利用者がフィッシングサイトであることを確認することでフィッシングサイトの共有を行うサイトである．そのため確認されている URL の中に正規のサイトが含まれている可能性がある [9]．今回は取得した URL はすべてフィッシングサイトであると仮定したうえでデータセットを作成した．フィッシングサイトのデータセットは Phishtank に掲載されている URL から Playwright ライブラリを使いウェブページへのアクセス試行し，アクセスすることのできた URL をデータセットとした．合計 1063 件の URL をデー

```
You are a web programmer and security expert tasked
with examining a web page to determine if it is a
phishing site or a legitimate site. To complete this
task, follow these sub-tasks:
1. Analyze the HTML, URL, and OCR-extracted text
screenshot image for any SE techniques often used in
phishing attacks. Point out any suspicious elements
found in the HTML, URL, or text.
2. Identify the brand name. If the HTML appears to
resemble a legitimate web page, verify if the URL
matches the legitimate domain name associated with the
brand, if known.
3. State your conclusion on whether the site is a
phishing site or a legitimate one, and explain your
reasoning. If there is insufficient evidence to make a
determination, answer "unknown".
4. Submit your findings as **pure JSON**:
with the following keys:
- phishing_score: int (risk score from 0 - 10)
- brands: str (brand name or None)
- phishing: boolean
- suspicious_domain: boolean
- reasoning: str (concise explanation of your conclusion)
Respond ONLY with the JSON object.
. Do not include explanation or markdown code block.
Website URL: \{url\}
HTML: \{html\}
OCR Text: \{ocr_text\}
```

図 1: フィッシング判定プロンプトの構造

```
Check whois registrar, org, creation_date,
and nameservers for anything suspicious.
```

図 2: Whois 情報

```
Check DNS, organization name, creation date,
and nameservers for anything suspicious.
```

図 3: DNS 情報

```
check the certificate information obtained to
see if there is anything suspicious.
Also, confirm that the contents of
the certificate are consistent with those of the predicted brand.
```

図 4: SSL 情報

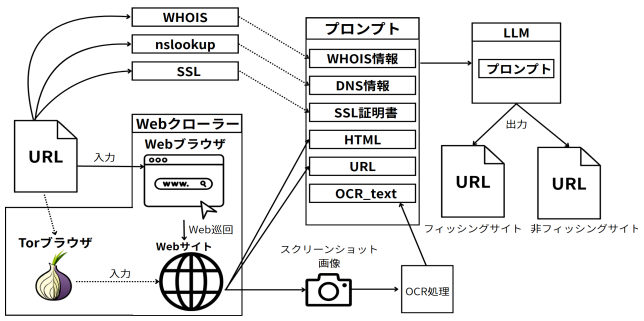

図 5: システムの概要

タセットとして用意した．しかし, 取得してから実験を行うまでに時間が経過したため消されてしまった URL や，LLM のトークン数の上限で出力を LLM からの返答を得ることのできなかったサイト，サイトが google のセーフサーチの対象になっており実験の時にはアクセスすることができなかった URL があるため，実際に実行することのできた URL は 1063 件よりも少ない数となってしまった．非フィッシングサイトは chatGPT にアメリカ，ロシア，日本，中国，イタリア，フランス，オランダ，カナダの 8 か国からそれぞれ 50 件の有名企業を出力させた. 出力させた有名企業のサイトを調べ，不審なサイトでないことを確認したのち URL を取得した．正規サイトの URL も Playwright ライブラリを使いウェブページへのアクセス試行し，アクセスすることのできた URL をデータセットとした．結果として 391 件の URL をデータセットとして用意した．正規サイトのデータセットもフィッシングサイトのデータセットと同様に実験の際に実行することのできた URL の数は 391 件よりも少ない数となっている．アクセスできない URL が多かったため，tor ブラウザを利用してはアクセスすることで入ることのできるサイトが多くなるかについても調べた．

3.2 閾値

出力された結果がフィッシングサイトと判断されたかそうでないかを判断する基準として phishingscore を利用する．閾値は出力されたフィッシングサイトと正規サイトの phishingscore をもとにヒストグラム**図 6** と**図 7** を作成し決定した．

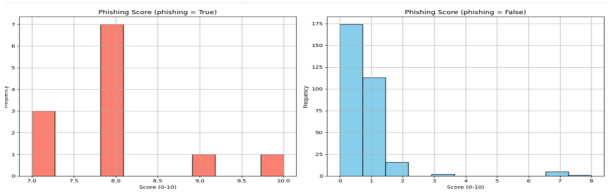

図 6: 正規サイトの phishingscore

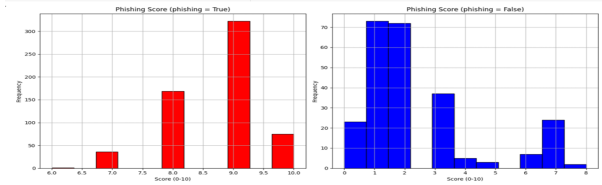

図 7: フィッシングサイトの phishingscore

図 6 図 7 のヒストグラムは URL，HTML，OCR を使用して実験を行った時のものである．ヒストグラムを確認するとどの実験でフィッシングサイトと判断されたサイトは phishingsore は 6 以上から数が多くなり，正規サイトであると判断された公式サイトは phishingsore は 3 以下のも

のが多かった．この図より phishingsore が 6 以上のものをフィッシングサイトと判断されたとして，それ以下のものは非フィッシングサイトと判断されたものとする．

3.3 評価指標

実験の評価指標として

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

を使用する．TP，TN，FP，FN はそれぞれ真陽性，真陰性，偽陽性，偽陰性を表す．official が official と判断された場合が TP の時と，phishing が phishing と判断された場合が TP の時の両方で計算を行う．accuracy の計算結果は official を TP としても phishing を TP としても結果は変わらないため，phishing の行にのみ記述する．フィッシングサイトと公式サイトのデータセットの数の差が多いため precision の数値が大きく変化することが考えられるので主に recall と accuracy に注目して評価する．

3.4 実験結果

① HTML+URL+スクリーンショット+OCR **図 8 表 1**
② ①+whois **図 9 表 2**
③ ①+nslookup **図 10 表 3**
④ ①+nslookup+tor **図 11 表 4**
⑤ ①+SSL 証明書**図 12 表 5**
⑥ ①+whois+nslookup+SSL 証明書**図 13 表 6**

①は正解が公式サイトのラベルであるものに関しては 97%で正確に判別することができている．しかしフィッシングサイトの精度に関しては 59%しかなく，高い精度とはいえない．

②は①と比較しても各箇所が数%異なるだけでほとんど同じ結果であると言える．

③は他の手法に比べ判別の精度が悪く①②と比較すると phishing の recall が約 5%程悪くなっているので精度が良いとは言えない．

④の手法では tor を利用している．③の結果と比較し手精度が大幅に良くなっていることが分かる．精度が良くなったと考えられる理由は公式サイトを用いたフィッシングサイトが他の手法では phishingscore が低く表示されていたサイトが tor を利用したことによりアクセスブロックされており画面にアクセスを拒否する旨のページが多くなってしまった．そのため phishingscore が危険なサイトであると判断し phishingscore が高いサイトが多くなったと考えられる．また正規サイトでも他の手法と比較すると誤判定の件数が増えているが，これも tor 利用によるアクセス拒否の画面で FN が増えてしまい公式サイトの recall が悪くなってしまったと考えられる．そのため精度が向上したわけではないと考えられる．

⑤の①+SSL 証明書の結果は phishing の recall が 57%と今回の実験の中で最も低くなった．理由として考えられることは google の正規サイトを用いたフィッシングサイトの判別率が悪かったからであると考えられる．whois では google の正規サイトを用いたフィッシングサイトの検出数が 311 件中 42 件だったのに対して，SSL 証明書では 321 件中 124 件と約 3 倍の数も誤判別された．

⑥の①+Whois+nslookup+SSL 証明書の結果は phishing の recall は 64%と⑤に次いで精度が悪かった．全体的に誤判別されたものは正規サイトを利用したサイトが多く，google を使用したものは 312 件中 81 件となった．正規サイトを用いたフィッシングサイトに関しては Whois+nslookup+SSL 証明書の情報はすべて正規サイトのものとなるので誤判定の割合が多くなると考えられる．

6 つの実験を通して tor を使った時以外はすべて official の recall は 90%を超えていた．このことから，正規サイトのラベルが付いたものから正規サイトを辺別する精度が高いことが分かる．

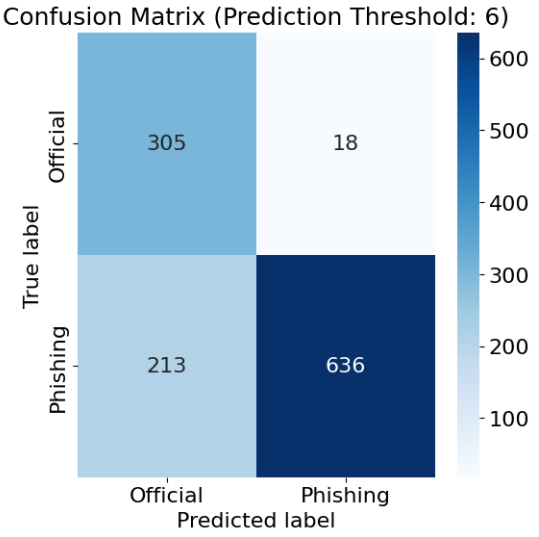

図 8: 基本の混同行列

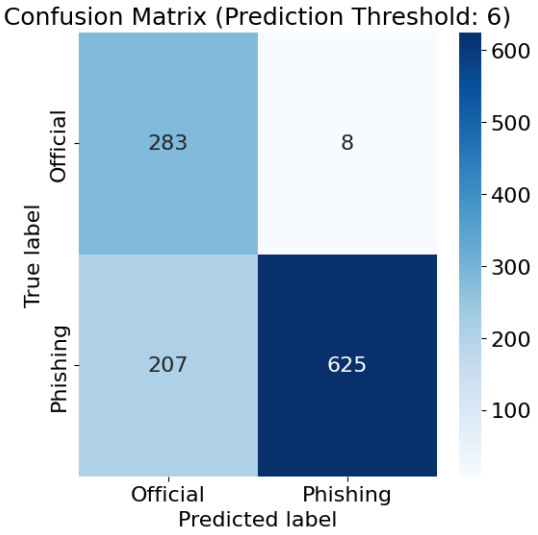

図 9: Whois の混同行列

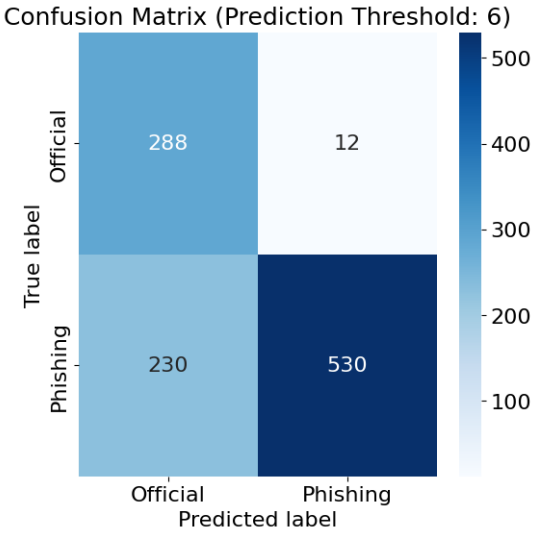

図 10: nslookup の混同行列

	precision	recall	f1-score	accuracy
official	0.59	0.94	0.73	
phishing	0.97	0.75	0.85	0.80

表 1: 基本の精度

	precision	recall	f1-score	accuracy
official	0.58	0.97	0.72	
phishing	0.99	0.75	0.85	0.81

表 2: Whois の精度

3.5 誤判定

誤判定をした理由には LLM が原因として挙げたものには以下のようなものがあった.

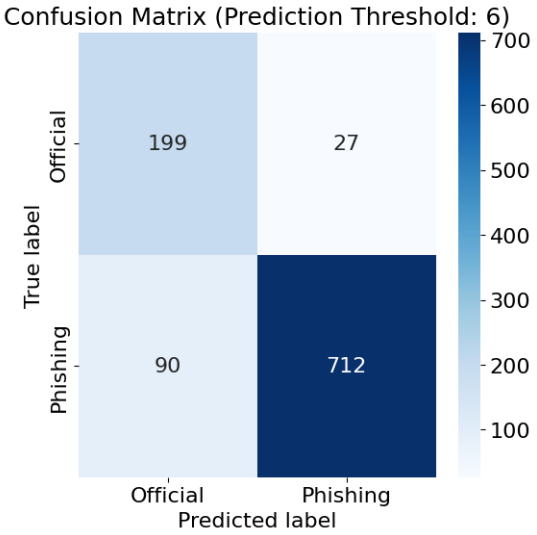

図 11: nslookup+tor の混同行列

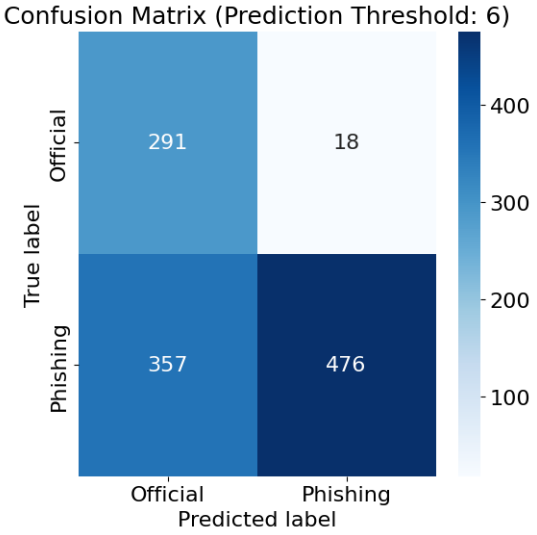

図 12: SSL 証明書の混同行列

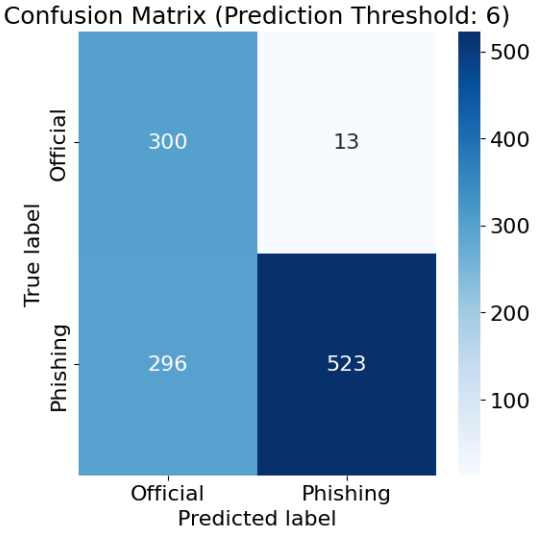

図 13: whois+nslookup+SSL 証明書の混同行列

	precision	recall	f1-score	accuracy
official	0.56	0.96	0.70	
phishing	0.98	0.70	0.81	0.77

表 3: nslookup の精度

	precision	recall	f1-score	accuracy
official	0.69	0.88	0.77	
phishing	0.96	0.89	0.92	0.89

表 4: nslookup+tor の精度

	precision	recall	f1-score	accuracy
official	0.49	0.94	0.61	
phishing	0.96	0.57	0.72	0.67

表 5: SSL 証明書の精度

	precision	recall	f1-score	accuracy
official	0.50	0.96	0.66	
phishing	0.98	0.64	0.77	0.73

表 6: Whois+nslookup+SSL 証明書の混同行列

3.5.1 フィッシングサイトが正規サイトに誤判断された要素と考察

フィッシングサイトの中には既存の正規サイトを使用したフィッシングサイトが存在する [11]. **図 14** のように本研究では google document や zoom を利用したサイトが見つかった. 正規サイトを利用しているため取得した Whois 登録情報, DNS 情報, SSL 証明書情報すべて正規のサイトのものが表示されてしまうため, フィッシングサイトであると判断されてしまう.

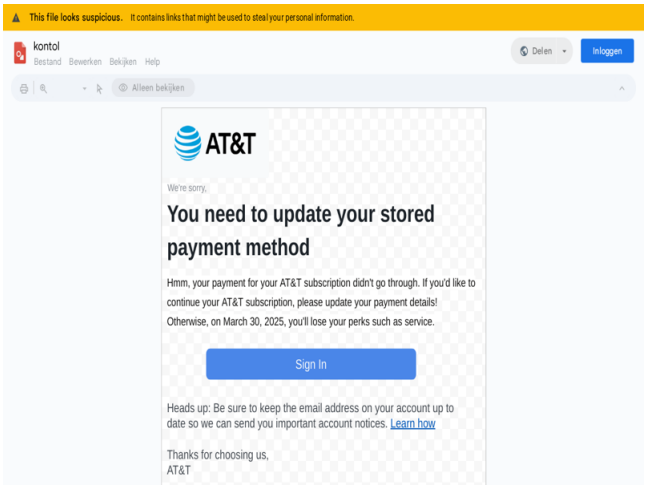

図 14: google document を利用したフィッシングサイト

①ドメインの一致

今回使用したデータセットには正規のサイトを利用したフィッシングサイトが存在した. このサイトでは Whois 情報, DNS 情報, SSL 証明書情報どれも元のサイトと同じ情報が表示されるので今回追加した情報からフィッシングサイトであると判断することができない. そのため以前からの HTML や OCR からの情報でしかフィッシングサイトであると判別するしかない. しかし正規の情報が混ざることで余分な情報となり LLM がフィッシングサイトでないと判断するケースが多く存在した.

図 16 のケースは google document を使用したフィッシングサイトであり, Whois 情報にもドメインにも不審な点はないと出力されている. しかし最後にタイトルが不自然で一般的な document 名ではなく OCR では Amazon とアカウント認証に関する言及があり, このページがフィッシング目的で使用されている可能性が示唆されていると出力されている. フィッシングサイトの可能性についても言及しているが, このフィッシングスコアは 2 であるためこのサイトはフィッシングサイトではないと判断している.

②短縮 url サービス

bitly や TinyUrl 等の短縮 URL サービスを利用したものでは, 短縮前の URL のが LLM に引き渡されてしまう. そのため Whois, DNS, SSL 証明書の情報も短縮前 URL の情報となる. 引き渡す HTLM に関してはリダイレクト後のコードを引き渡すが, Whois 等の情報により安全なサイトであると誤判断される結果があった. また短縮後の URL が削除されていた場合画面に表示されるのは短縮 URL サービス正規のサイトで url Terminated など URL が削除された旨の内容が表示された. そのためサイトとしては安全なため, フィッシングサイトでないという誤判断が多く見られた.

③ e コマースサイト

e コマースサイトではしばしば詐欺業者が存在するサイトそのものは安全なサイトであってもサイトで販売をしている業者が詐欺を行っているのであればシステム的には安全なサイトである. よってこの手法では見つけることはできず安全なサイトであると判断したのではないかと考えられる [12].

④リダイレクト

最終的には正規のサイトにアクセスするが, アクセスする途中不必要なサイトを経由することでマルウェアのダウンロードをさせる rive-by Download 攻撃という手法が存在する [10]. URL には最終的にたどり着くサイトの URL の含まれているため有名な URL を使い一見無害に見える URL にすることでユーザーは安心を促して接続させる. この手法のように最終的に接続する先が一般的に知られている URL であった場合安全なサイトであると判断をしてフィッシングサイトではないと判断をしてしまうと考えら

れる [13]．**図 15** のように最終的に wikipedia に行きつくような URL が本研究では見られた．

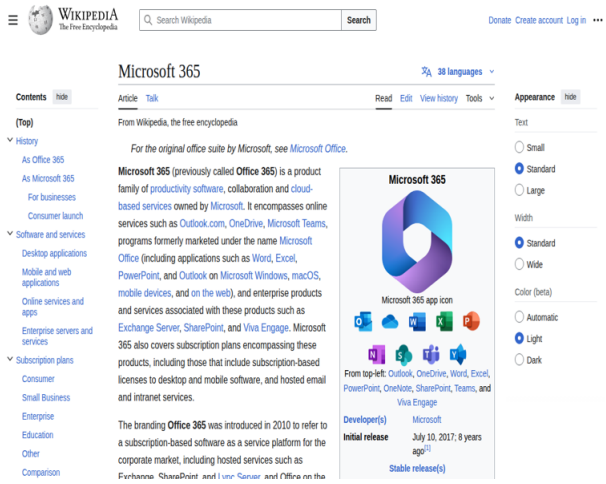

図 15: リダイレクトされ最終的に到達したページ

3.5.2 正規サイトがフィッシングサイトに誤判断された要素と考察

誤判定をした理由には LLM が原因として挙げたものには以下のようなものがあった．

①アクセス制限

bot でのアクセスに対して制限がかかっていたサイトでは画面の OCR でアクセス制限があるため，不審なサイトである可能性があると出力されたものが多くあった．また証明書情報にエラーがあり情報を複合的に推測するとフィッシングサイトの可能性が高いと判断されたものもあった．

② captcha

アクセスしたサイトには captcha が使われているサイトがあった．人間かボットかを判断する画面は一般的な正規サイトでは正常な動作であるが whois 情報が記載されておらず，追加の文脈がないため判断できないのでフィッシングサイトと分類されたものや，Whois 情報は正確であるが悪意のある captcha の可能性があるためフィッシングサイトと分類されたものもあった．

4. 終わりに

本研究では Whois を用いたときがフィッシングサイトに recall が最も高くなり，SSL 証明書を用いたときが最も recall が低くなることが分かった．しかし Whois，DNS 情報，SSL 証明書のどの要素がフィッシングサイトの検出に効果的かを判別することができなかった．本研究で用いたデータセットには正規サイトを利用したフィッシングサイトが多くあり Whois，nslookup，SSL 証明書の情報はより精度を低くしてしまう原因になると考えられる．正規サイトを用いたフィッシングサイトを LLM を用いて判別する手法に関しては今後の課題となると考える．また今回使用し

> The URL is a Google Docs drawing, and the HTML indicates it is a legitimate Google service. The WHOIS and certificate information for the domain google.com are valid and consistent with a legitimate Google domain. However, the title "LKWEPIJWHNJOITJOS" is odd and not consistent with a typical document name, and the OCR text mentions Amazon and account verification, suggesting this page might be used for phishing purposes.

図 16: 誤判定例

たは LLM のモデルは軽量なモデルである gemini2.0-flash を用いたが，高度な推論を行うことのできる別のモデルを使うことで結果がどれだけ変わるかについても研究の余地がある．本研究で得られた結果は，本手法の適用可能性や改良の余地を示しており，今後の発展に向けた貢献を有している．

参考文献

[1] Tang, S. et al.: Clues in Tweets: Twitter-Guided Discovery and Analysis of SMS Spam, Proc. ACM CCS 2022.

[2] Szurdi, J. et al.: Where are you taking me? Understanding Abusive Traffic Distribution Systems, Proc. WWW ' 21.

[3] Koide, T. et al.: It Never Rains but It Pours: Analyzing and Detecting Fake Removal Information Advertisement Sites, Proc. DIMVA 2020.

[4] Jie Zhang, et al. "When LLMs Meet Cybersecurity: A Systematic Literature Review", arXiv:2405.03644, 2024

[5] Takashi Koide, et al. "Detecting Phishing Sites Using ChatGPT", arXiv:2306.05816v2, 2024

[6] 吉田 蓮，和泉 諭，菅沼 拓夫．マルチモーダルな情報を活用した LLM によるフィッシングサイト検知手法の提案．IOTS2024 .2024/12/6

[7] Hiroki Nakano,Takashi Koide, Daiki Chiba.Autonomous Scam Website Detection Using Large Language Models. Computer Security Symposium 2024. 22-25 October 2024

[8] Takashi Koide,Hiroki Nakano,Daiki Chiba.Leveraging Large Language Models for Phishing Site Detection.Computer Security Symposium 2024.22-25 October 2024

[9] Yuichi Nakamoto, Masahiko Katoh.The points to note when making datasets for phishing websitesdetection.Computer Security Symposium 2021.26 - 29 October 2021

[10] Koya Ozaki,Shinya Ueyama,Tatsuya Konishi,Masato Yamazaki, Tsubasa Bando,Takashi Kobayashi1."Proposal to support analysis of Drive-by Download attack by highlighting malicious URL",Computer Security Symposium 2017. 23 - 25 October 2017

[11] Sayak Saha Roy,Unique Karanjit,Shirin Nilizadeh.A Large-Scale Analysis of Phishing Websites Hosted on Free Web Hosting Domains.arXiv:2212.02563v2 [cs.CR] .10 Jan 2024

[12] Vinicius Facco Rodrigues, Lucas Micol Policarpo, Diórgenes Eugênio da Silveira, Rodrigo da Rosa Righi, Cristiano André da Costa, Jorge Luis Victória Barbosa, Rodolfo Stoffel Antunes, Rodrigo Scorsatto, Tanuj Arcot.Show more,Fraud detection and prevention in e-commerce: A systematic literature review.Volume 56,

November – December 2022, 101207
[13] Soheil Khodayari , Kai Glauber , Giancarlo Pellegrino.Do (Not) Follow the White Rabbit: Challenging the Myth of Harmless Open Redirection.Network and Distributed System Security (NDSS) Symposium 2025 .24-28 February 2025, San Diego, CA, USA