東京大学
情報理工学系研究科　電子情報学専攻
修士論文

顔画像生成技術を用いた偽造ウェブサイト判別支援手法の設計と実装

Facial Image Generation-based Method
for Supporting User Efforts to Identify Fraudulent Websites

48-206456
山崎　慎治
Shinji Yamazaki

指導教員　宮本　大輔　准教授

2022 年 1 月

概要

情報技術の普及に伴って，情報技術を悪用するサイバー攻撃が社会的問題になっている．なかでもフィッシング攻撃は，ソーシャルエンジニアリングの手口を利用し機密情報を詐取する詐欺行為であり，サイバー社会に対する脅威の一つである．さらに，フィッシング攻撃は多様性を増し手口も洗練され，被害は年々拡大している．これまでに，システムによるフィッシング攻撃の検知，ユーザーに対する教育による被害の防止，ユーザーがフィッシング攻撃と判断できるような意思決定の支援などの，多角的な手法の研究により数多くの対策が提案されている．ここで，フィッシング攻撃の本質はユーザーを騙すことであり，この特徴はフィッシング攻撃が最初に報告された 1995 年から現在に至るまで不変である．そのため，ユーザーがフィッシングサイトを判別し正しく意思決定できるように支援する手法は重要である．しかし，SSL 証明書の警告表示やドメイン名の強調表示のような従来の手法は一般のユーザーにとって難解である，という問題が先行研究で明らかになっている．そこで本研究では，ユーザーに提示する簡易な情報の形態として顔画像を用い，フィッシング攻撃の判別を支援するシステム NOPHICE を提案し実装した．そして提案システムの有効性を検証するために，アンケート調査を実施した．結果として，提案システムはフィッシング攻撃に対して利便性と安全性を向上させることがわかった．特に，ユーザーがフィッシングサイトを正規のウェブサイトと誤認してしまう割合を有意に減少させた．さらに提案システムの応用やアクセシビリティ，システムを普及させるときの課題点，システムに対する攻撃などを検討し議論した．

Abstract

Cyberattacks have become a social problem with the spread of information technology. Among them, phishing attacks are one of the most severe threats to cyber society as it is a fraudulent act of stealing confidential information by using social engineering techniques. Moreover, phishing attacks are becoming more diverse and sophisticated, and the number of victims is increasing year by year. Numerous anti-phishing measures have been proposed through the work of multifaceted methods, such as system-based detection of phishing attacks, user education to prevent attacks, and decision-making support to help users identify phishing attacks. Here, the essence of a phishing attack is to deceive users, and this characteristic has not changed since the first report of a phishing attack in 1995. Therefore, it is essential to support users to identify phishing sites and make correct decisions. However, conventional methods such as SSL certificate warning and domain name highlighting are complicated for ordinary users. However, conventional methods such as SSL certificate indicators and domain name highlighting are complex for general users to understand. In this study, we proposed and implemented NOPHICE. This system helps users identify phishing attacks by using facial images as a simple form of the information shown to the user. We then conducted a questionnaire survey in order to verify the effectiveness of the proposed system. As a result, we found that the proposed system improves convenience and security against phishing attacks. In particular, it significantly reduced the rate users misidentified phishing sites as legitimate websites. In addition, we examined and discussed the application of the proposed system, accessibility, challenges in spreading the system, and attacks against the system.

目次

図目次

表目次

第 1 章

序論

1.1　背景

コンピュータやインターネットに代表される情報技術（Information Technology, IT）は，登場以来人類の生活に驚くべきほど浸透し，生活基盤の一つとして世界の発展に寄与している．

一方で情報技術の悪用しコンピュータや情報システムに攻撃する行為はサイバー攻撃と形容され，人々の生活を脅かす存在として問題となっている．近年発生したサイバー攻撃の事例として次に挙げる：日本の暗号資産取引所から当時の換算レートとして 580 億円相当の暗号資産が流出した [8]．著名人のプライベート写真がクラウドプラットフォームから大量に流出しインターネット掲示板に掲載された [9]．アメリカ合衆国最大の石油パイプラインがランサムウェアによって操業停止となり，身代金として 440 万ドルを支払った [10]．国家の支援するサイバー攻撃グループによって特定の国家に対するスパイ行為や介入行為を行い，たとえばアメリカ合衆国大統領選挙に対してロシア連邦の支援したグループが世論誘導を行ったと指摘する国際政治学の専門家が存在する [11]．

こうした状況から，2011 年 7 月にアメリカ合衆国国防総省はサイバー空間を陸・海・空・宇宙空間に次ぐ第五の戦場として，国家安全保障の立場からサイバー空間やサイバー攻撃を再定義するに至っている [12, 13]．同様に，ロシア，中国，欧州各国を始めとする世界中の国々で，サイバー攻撃を国家安全保障上の課題として捉え，様々な動きを見せている [14]．

このサイバー攻撃の一種にフィッシング攻撃がある．フィッシングとは，ソーシャルエンジニアリングの手口によって，個人情報や金融口座の情報といった個人や団体の保有する機密情報を開示するように誘導し，それらの情報を入手，或いは利用する詐欺行為である [15, 16]．ここで，ソーシャルエンジニアリングとは，攻撃対象の保有する機密情報を意図せずに漏洩させる様に仕向けるための心理学的操作手法を表す [17]．フィッシングは 1995 年の America Online (AOL) における事例の報告 [18] から始まり，20 年以上経った現在においても尚，脅威となっている．

フィッシングは，典型的に次のような手順によって攻撃が展開される．(1) 準備段階として攻撃者は攻撃対象となるユーザーのメールアドレスを収集し，誘導先となる偽のウェブサイトを用意する．このとき，偽のウェブサイトは攻撃者が用意したサーバーにホストされる他，インターネット上に存在する脆弱なサーバーを乗っ取った上で展開されることもある．(2) 続いて，送信者を偽ったフィッシングメールを，攻撃対象のユーザーに送付する．フィッシングメール中に記載された偽サイトへのリンクをクリックするように誘導する．(3) 予め設置した偽のウェブサイトに誘導したユーザーに個人情報を入力するように更に誘導する．(4) 入力された情報を入手し悪用する．悪用例として，入手した個人情報を収集し販売する，入手した情報を元に金融口座などにアクセスし金銭を得る，入手した情報を用いて更にフィッシング攻撃を展開することが挙げられる．

しかし，近年では利用するメディアや技術的なアプローチが様々になり，フィッシング行為は多様化が進んでいる．たとえば，ショートメッセージサービス（SMS）を利用したスミッシング（Smishing, SMShing)，インスタントメッセンジャーを利用する IM Phishing, 電話システムを利用して情報を盗む Vishing (Voice Phishing）などがある [19]．特に近年ではスミッシングの脅威が増加しており [20]，一部の携帯電話事業者は利用者にスミッシングに特化した対策サービスを提供する計画を発表している [21, 22]．

また，フィッシング攻撃は洗練され巧妙化している．一つに，機械翻訳技術の向上がある．たとえば，一般に日本語は習得が難しい言語とされており，日本語を母語としない攻撃者が日本語でフィッシング攻撃を展開することは容易でなかった．しかし，自然言語翻訳技術の発達によって誰もが自然な日本語を生成できるようになり，フィッシング攻撃から不自然な日本語が見当たらなくなりつつある．他にも，フィッシングキットの存在によって多様な人が高度のフィッシング攻撃を手軽に展開できること，フィッシングサイトの HTTPS 対応 [15] により SSL 証明書の有無ではフィッシング攻撃を判別できないことが挙げられる．

こうしたフィッシングへの対策に関する既存の研究は，大きく 3 つの方向性を持って為されていると考えられる．第一には，フィッシングそのものを対象として機械的に処理する，フィッシング行為の検知および分類の研究がある．第二には，フィッシングの被害者となり得るユーザーへの教育による被害の防止を狙った研究があり，そして第三にユーザーがフィッシング攻撃に直面した時に，フィッシング攻撃と判断できるように意思決定を支援するための方法を模索する研究がある．

フィッシング攻撃はソーシャルエンジニアリングの手法を利用する攻撃であり，その本質は人を騙す行為にある．フィッシング攻撃が洗練され手口が多様化する中でも，情報技術を利用するユーザーを欺くことで攻撃が成立することに変わりはない．そのため，フィッシング対策の研究もユーザーに着目し続けなければならない．つまり，ユーザーがフィッシング攻撃に騙されないように意思決定を支援する研究の重要性が増している．

図 1　運送業者を騙ったスミッシングの例

1.2　目的

これまでにユーザーの意思決定を支援する研究は，10 年以上に渡って数多くなされてきた．しかし，フィッシングの被害は依然として増加しており，これまでの研究で効果的な手法を打ち出すことができていなかったと言わざるを得ない．なぜなら，フィッシングの被害に遭うような一般的なユーザーと，研究の中で想定しているユーザー像とで，情報技術への理解度の乖離があるからだと考えられる．これは，研究者にとっては提案手法が意図した通りの効果を発揮できない，ユーザーにとっては活用が難しすぎるという問題に繋がる．

そこで本研究では，一般的なユーザーであっても十分に活用可能な平易な方法であることを条件に，ユーザーの意思決定を支援する手法に取り組む．ユーザーの意思決定を支援するにあたって，情報技術の知識を必要としない形式のインジケーターを利用したシステムを提案し，実装することを目的とする．さらに実装したシステムをアンケート調査を通じて評価し，発展させるために議論することも本研究の目的である．

1.3　構成

本論文は修士研究を総括するにあたって，次のように構成される．2 章では，ユーザーの意思決定を支援する関連研究を概観する．また，認知心理学の観点から顔画像と記憶の関係についてもまとめる．3 章では，顔画像をインジケーターに用いた NOPHICE システムを提案し，実装する．4 章では，ユーザー調査を通じて NOPHICE システムを評価し，5 章では，NOPHICE システムを多角的に議論する．最後に 6 章で本研究のまとめとする．

第 2 章

関連研究

2.1 ユーザーの意思決定を支援する手法についての既存研究

本節では，フィッシング攻撃に直面したユーザーが正しく意思決定をした上で攻撃を回避できるように，警告の改善や情報の提示などの手法によってユーザーを支援することを目的としたユーザーインターフェースの既存の研究を取り上げる．

フィッシング攻撃に直面したユーザーは，攻撃の展開に合わせて大きく 2 つのタイミングで意思決定を迫られる．すなわち，電子メールに記載されたリンクをクリックしてウェブページにアクセスするタイミングと，ウェブページにアクセスした後にフォームに情報を入力するタイミングである．既存の研究はそれぞれのタイミングに着目し，前者であれば電子メールクライアントを拡張することで，後者であればウェブブラウザを拡張することで，様々な手法を提案してきた．

2.1.1 電子メールクライアント

Nicholson ら [1] は，電子メールの送信に関する情報を強調表示し，受信に関する情報を追加で提示する画面表示によって，フィッシングメールに対する感度を向上させようと試みた．強調される送信に関する情報とは，具体的には送信者名，メールアドレス，受信時刻の 3 点であり，正規のメールであれば営業時間として一般的な時間に送付されるべきであろうという推定に基づいている．また，追加で提示される受信に関する情報とは，他の同僚が同様の電子メールを受け取った割合であり，正規のメールであれば特定の宛先に送信されて無差別には送付されないであろうという推定に基づいている．フィッシングメールと無害なメールとを見分ける実験の結果から，送信に関する情報の強調表示には効果があったと報告している．また，ユーザーの判別結果の偏りは提案法による表示によって変化しなかったことから，表示の内容に基づいてユーザーが判断したと主張している．

一方で Volkamer ら [2] は，電子メールクライアントソフトの拡張機能として TORPEDO を提案した．TORPEDO は，ユーザーが電子メールに含まれるリンクをクリックしようとす

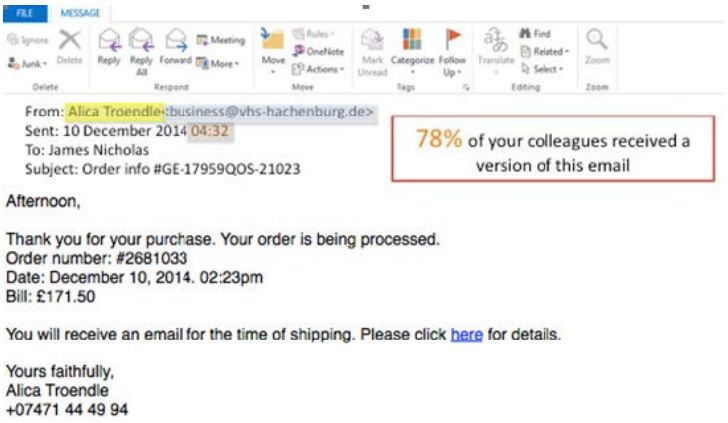


図 2　Nicholson ら [1] の実装図．送信者と送信時刻をハイライトし，同様のメールを受信した同僚の割合を右上に表示している

図 3　Volkamer ら [2] が提案したツールチップ．カーソルを合わせたリンクのリダイレクト先アドレスを表示している

るとツールチップが浮かび上がり，リンクが有効化されるまでに遅延を設ける．TORPEDO では大きく 3 つの機能をツールチップ内で提供している：(1) リンク先のアドレスを強調して表示し，注視するよう文章で促す．(2) リンク先のドメイン名は文字間隔を広げて表示することで誤読を防ぐ．例えば，“m” と “rn” の判別が容易になる．(3) ツールチップ内のボタンをクリックすることで，件名や本文中に挿入されたブランドロゴマークなどの他の要素に惑わされないように促したり，フィッシングサイトのアドレスとして一般的に用いられる手法を紹介したりする解説文を表示する．Volkamer らは，直ちにその場でツールチップを表示する TORPEDO が，ユーザーの判別にどの程度効果があるか，電子メールクライアントソフトで一般的なステータスバーの URL 表示と比較した．実験の結果から，提案法を利用したユーザーは有意により正しく判別できたと報告している．

Petelka ら [3] は，既存の電子メールクライアントソフトの警告が具体的ではないとして，次の 3 つの方法で警告の改善を提案した：(1) フィッシング攻撃の警告を電子メール本文

図 4　Petelka ら [3] が提案した警告の改善．メール本文中のリンクが無効化されており，アクセスするにはポップアップされた警告画面内に記載されたリンクをクリックしなければならない

中の疑わしいリンクの近くに表示する．(2) 元の本文中のリンクを無効にして，表示された警告の領域にある URL をクリックさせる．(3) リンクにカーソルを合わせたときに警告をポップアップして表示する．Petelka らは，701 名に対するオンラインの調査結果から，従来のバナーによる警告よりもリンクの近くに表示された警告のほうがよいこと，元のリンクを無効にする方法は有意に効果があったこと，警告のポップアップ表示には有意差が見られなかったことを報告している．

2.1.2　ウェブブラウザ

2.1.2.1　ツールバー

ツールバーとは，ユーザーのインストールによってウェブブラウザの機能を拡張する，帯状のインターフェースである．ツールバーはブラウザの開発元以外の第三者でも提供できる特徴があり，広く利用されてきた．しかし，現在のウェブブラウザのユーザーインターフェースではツールバーの領域が削除されており，併せて主要なツールバーのサービスも終了している [23, 24]．

Wu ら [25] は，このツールバーを用いたセキュリティ機構が実際にユーザーをフィッシングから保護しているのかを検証するために，実験環境下で被験者にメールが送られ，その一部がシミュレートされたフィッシングサイトに誘導するものである，というシナリオでユーザー調査を行った．ここで多くのユーザーは，ウェブサイトのコンテンツが正規のウェブサイトのものと捉えると，ツールバーの警告を見過ごしてしまうことが報告されている．また，

正規のウェブサイトがログインページを HTTPS 対応させないことで，ユーザーがフィッシングであるかどうかを判別することが難しくなっていることを指摘している．他にも，ツールバー上の消極的な警告ではなくポップアップ画面のような積極的に割り込む警告のほうが効果的だとも指摘している．

2.1.2.2 警告画面

頻発するセキュリティ警告画面には，馴化によってユーザーは警告を信頼しなくなり，効果的に被害を防止できなくなる問題がある．この問題に対して Bravo-Lillo ら [4] は，最も重要な情報に注意を向けるような UI としてアトラクターを設計し，これを意思決定を支援する手法として提案した．馴化の下での実験や馴化そのものの影響を調査した結果，表示された重要な情報をユーザー自身が入力する方式（“Type”）やユーザーがマウスポインタでなぞる方式（“Swipe”）の効果が高いことが報告されている．また一連の調査を踏まえ，馴化の影響を受けた中でのユーザーテストを実施することを提案している．

2.1.2.3 アドレスバー

多くのウェブブラウザは上部に配置したアドレスバーに，ユーザーに表示しているウェブページの情報を表示している．アドレスバーに表示する情報として，ウェブページのアドレスである URL と SSL 証明書インジケーターがある．

Xiong ら [26] は，アドレスバーに表示している URL のうちドメイン名を強調表示するウェブブラウザの実装が，フィッシング攻撃への対策として実際に効果があるか検証した．アイトラッキングを用いた実験では，ユーザーはアドレスバーではなくウェブサイトの内容に注意してウェブサイトの悪性を判断していることがわかった．また，アドレスバーを注視するよう被験者に指示を出したあとでは，ドメイン名が強調されていない場合では SSL 証明書のインジケーターに着目し，ドメイン名が強調されているとドメイン名にも着目するようになる，という振る舞いの差異が観測された．一方で，ドメイン名の強調の有無によってウェブサイトの判別精度は有意に変化しなかった．Xiong らは，ユーザーがウェブサイトの判別に正しく活用するだけの，アドレスに関する知識を持ち合わせていなかったことを指摘している．実際に，アイトラッキングを用いた実験後のアンケート調査では，ほとんどの被験者はウェブサイトの見た目で判断したと回答している．アドレスバーを見て判断した一部のユーザーであっても，SSL 証明書のインジケーターや HTTPS 通信であるかどうか，URL の長さを判断の根拠としており，ドメイン名の強調表示に言及している被験者はいなかった．中には，URL が長いほどウェブサイトが安全ではないという間違った知識を持ち合わせたユーザーが存在することも，アンケート調査の結果から判明している．

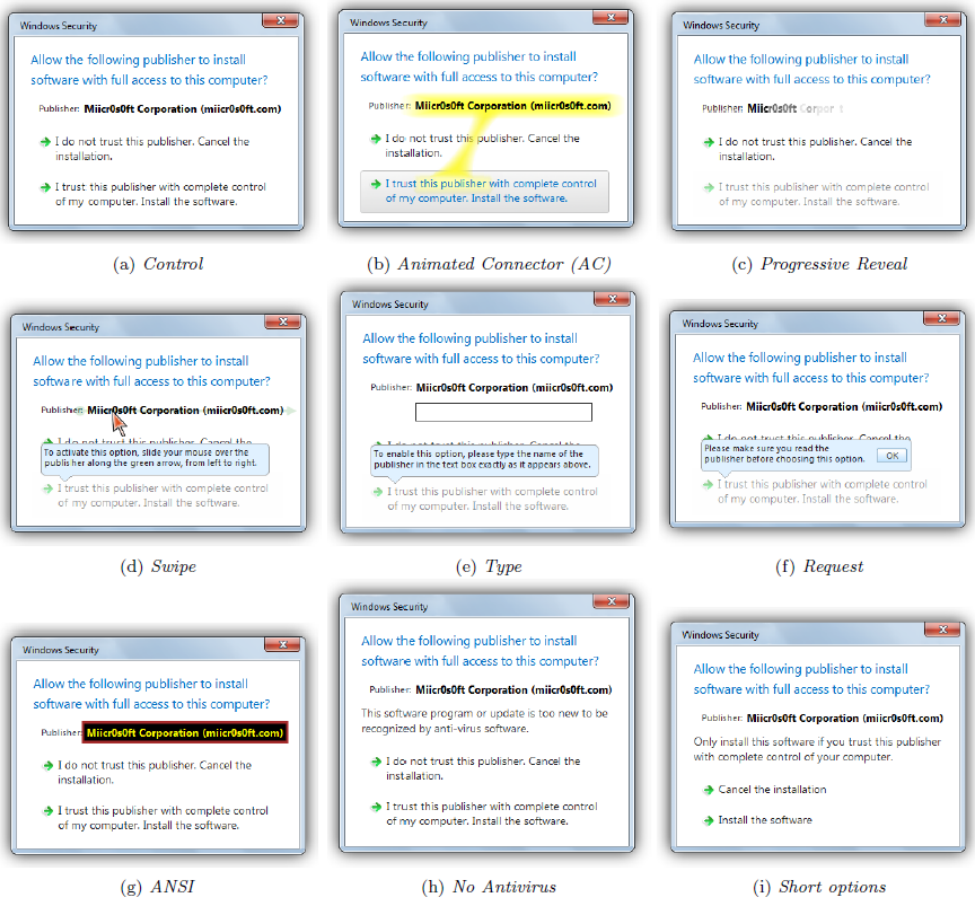


図 5　Bravo-Lillo ら [4] の実験で用いられたアトラクターの例．不審な発行元の情報が強調表示されている

一方で，SSL 証明書はウェブサイトのアイデンティティを示し，正しく使用すれば正規のウェブサイトかフィッシングサイトかを判別するための鍵となる要素である．ウェブブラウザには SSL 証明書に記載された情報をユーザーに提示する UI がある．

Felt ら [5] は，ブラウザが不審な SSL 証明書を検知した時に表示する警告を改善することで，よりユーザーが危険性を理解し安全な選択するような UI を目指した．警告を簡潔かつ具体的かつ技術的用語を排することで改良した結果，ユーザーはより安全な選択をしフィッシングを回避したが，警告が意味する SSL 証明書についての危険性の理解を向上させられなかった．

また Felt ら [6] は，アドレスバーの横に表示される接続の安全性を示すアイコンであるインジケーターについての，多様な形状や色を比較するサーベイ調査を行うことにより，インジケーターの改良を目指した．調査結果を元に，状態によって色や形状がはっきりと変化し，

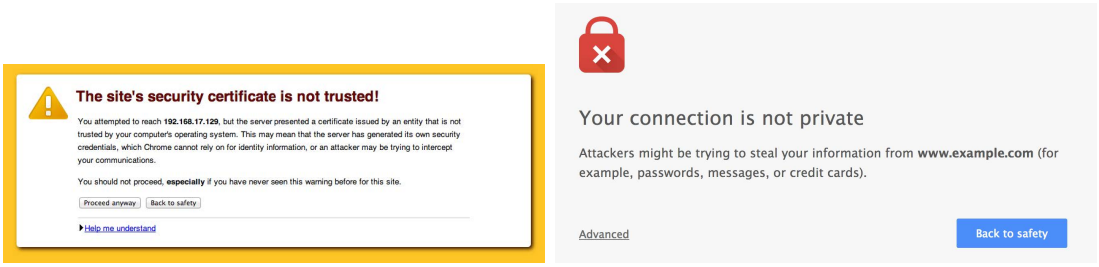


図 6　従来の警告画面（左図）と Felt ら [5] が提案した警告画面（右図）

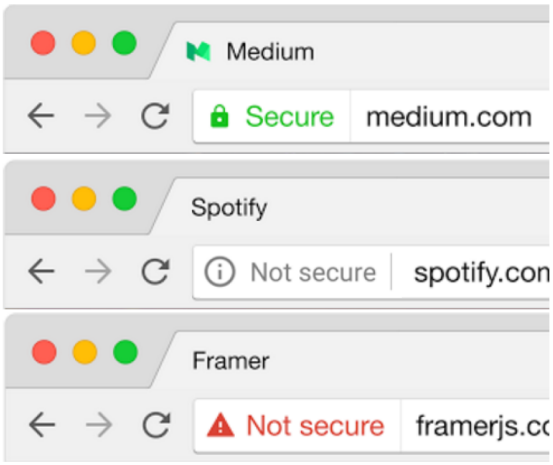


図 7　Felt ら [6] が提案したインジケーターのセット

簡潔に説明する単語を付随させたインジケーターのセットを提案した．この研究では，インジケーターを設計する際に，スマートフォンなどの小さなデバイスでも表示できること，色覚特性の影響を受けないように形状のみで理解可能であること，アイコンの意味を推定できないユーザーを考慮して言語で説明可能であることが要請されているとも指摘している．

ここで Thompson ら [27] は，EV-SSL 証明書の緑色を用いて強調表示するブラウザの UI がどれだけユーザーに影響を及ぼしているか調査した．インジケーターは改良されたもののユーザーの SSL 証明書に対する理解に影響を及ぼさず，EV-SSL 証明書の緑色の強調表示を無効化しても影響がなかったことを報告している．

2.2　ユーザーの意思決定を支援する既存研究の整理

前節では，ユーザーがフィッシング攻撃を回避できるように，ユーザーの意思決定を支援する多様な研究が行われてきたことをまとめた．しかし，現在では十全に活用できないアプローチの研究がある．例えば，電子メールクライアントでの手法はそのままではスミッシ

ングに対して適用できない．電子メールで定義されている SPF や DKIM のような機能は SMS に対して定義されておらず，一般的な SMS クライアントは簡潔な実装になることが多いからである．ウェブブラウザのツールバーについても，2022 年現在で広く利用されるウェブブラウザでは，ツールバー領域がインターフェースから削除され活用できない．

また Felt ら [5] の研究や Thompson ら [27] らの研究では，ユーザーの理解を得ることの難しさに触れている．情報技術に精通したユーザーであれば，URL や SSL 証明書などのウェブの知識，SPF や DKIM に代表されるような電子メールのセキュリティ技術の知識，攻撃者が利用する攻撃手法や動向といった情報を組み合わせて，フィッシング攻撃を判別している．フィッシング対策協議会の発行するパンフレット [28] でも，洗練されたユーザーが活用している判断基準を一般のユーザーも活用できるように説明することで，対策方法を啓蒙している．

特にフィッシング対策協議会は正しい URL にアクセスすることをフィッシング対策の第一に掲げている．しかし，一般のユーザーにとって URL を判別することは難しい．URL の持つ階層構造やトップレベルドメイン，ドメイン名といった特徴についての知識が常識として流布しているかは疑問の余地がある．実際，Albakry ら [29] の調査によれば，ユーザーは URL 内に含まれるブランド名の文字列を信用する傾向があり，階層構造を理解して判断するユーザーは情報技術の利用時間に関わらず少ないことが報告されている．Albakry らは，こうしたユーザーの短絡的な判断は多くの場面で労力を削減することから，ユーザーの傾向を批判することはできないとしている．しかし，ユーザーが URL の規則を理解していない以上，ユーザーが正規のウェブサイトの URL とフィッシングサイトの URL の判別を行うことは容易ではない．

そこで，URL を文字列以外の形式で表現することで，フィッシングサイトと正規のウェブサイトとの判別を支援できる可能性がある．機械的な検知ができなかったフィッシングサイトであっても，正規のウェブサイトと違うウェブサイトであることにユーザーが気づくことのできる形式であり，URL の規則および知識を必要としない方法が望ましい．

2.3 認知心理学の観点から考える顔画像の利点

2.3.1 記憶と想起

記憶は記銘・保持・想起の 3 つの過程から構成されていると考えられている．中でも想起は保持されている情報を適切に思い出す過程を指す．想起を測定する方法として，古くから再生課題（recall）と再認課題（recognition）が用いられている．再生課題は，保持している情報を想起して生成する課題であり，手法によって自由再生（free recall）・手がかり再生

（cued recall）・系列再生（serial recall）に分類される．また再認課題は，提示された情報が記憶として保持されているものと一致するか参照する課題であり，手法によって項目再認（item recognition）・連想再認（associative recognition）などに分類される．[30]

2.3.2 画像優位性効果

画像による刺激は言語による刺激よりも記憶に残りやすく，この現象を画像優位性効果（Picture Superiority Effect, PSE）と呼ぶ．画像優位性効果は，1970 年頃から認知心理学の分野で主張され，現在に至るまで研究対象となっている．画像優位性効果は，実際に複数の研究によって認められている．例えば，自由再生課題 [31,32]，系列再生課題 [33]，項目再認課題 [34–36]，連想再認課題 [37] のそれぞれで効果を示す研究が多数報告されている．

画像優位性効果を説明する理論として，主に 2 つの仮説が提唱されている．第一に二重符号化理論（Dual Coding Theory, DCT）[38] がある．画像刺激は画像コードと言語コードの 2 つの独立した記憶経路で記憶されることに対して，言語刺激は言語コードのみの経路で記憶される．更に画像コードと言語コードは相互に参照されるため，二重に符号化される画像刺激は想起されやすくなるとするものである．

第二に，示差性仮説（Distinctiveness Theory）がある．示差性仮説は，言語に比べて画像の方が識別性が高いとするもので，感覚・知覚的な識別性に着目した物理的示差性仮説と概念的な識別性に着目した概念的示差性仮説に大別される．物理的示差性仮説では，画像刺激は言語刺激に比べて物理的特徴の変化が大きいために画像優位性効果が生まれるとする．概念的示差性仮説では，画像の識別は単語の識別よりも認知的プロセスとして深いレベルの記憶処理を引き起こすために画像優位性効果が生まれるとする [39].

ただし，画像優位性効果の理論を実験で解明することは難しい．実験で各理論の影響の有無を検証するために変数を一つだけ操作しようとしても，記憶の過程で他の変数に影響を及ぼしてしまい，一つの変数を操作しつつ他の変数を一定に保つ実験が容易でないからである．また，被験者の画像記憶能力が非常に高く，画像刺激では天井効果が常に懸念されることも挙げられる [40]．しかし，画像優位性効果の存在は実験で認められており，また，画像の記憶は高水準で数カ月に渡って持続することも報告されている [41].

ヒトが記憶できる画像の容量も大きく，そして細部まで記憶していることがわかっている．Brady ら [7] は，被験者に 2,500 の画像を記憶してもらった後に 3 つの再認課題に取り組んでもらう実験を行った．全く異なった画像を比較した再認課題（“Novel”）では 92% の正答率を示した．さらに同カテゴリーの画像を比較した再認課題（“Exemplar”）では 88% の，向きや配置が変化しただけの画像を比較した再認課題（“State”）では 87% の，軽微な差異にもかかわらず高い正解率を示した．

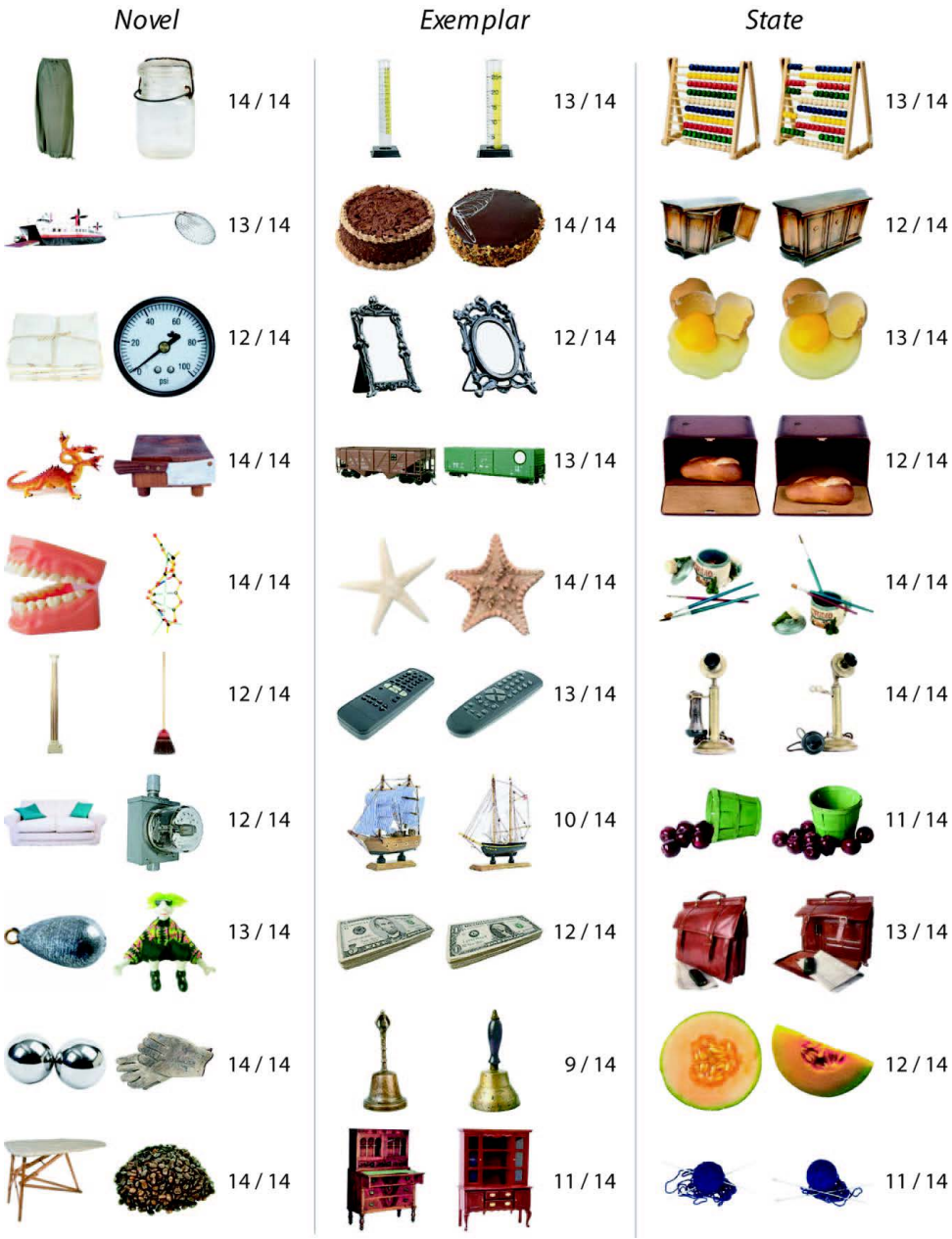


図 8　Brady ら [7] の実験で用いられた再認課題の一部と正答率

2.3.3　顔画像と記憶および知覚

画像は記憶することが容易であることは前項で述べた通りであるが，中でも顔画像は記憶することと知覚することに利点がある．

Isola ら [42] の調査によれば，様々な画像を比較すると，閉鎖空間で顔が写った人の画像は記憶に残りやすく，風景画像は記憶に残りにくいことが報告されている．また，Jenkins

ら [43] は，顕在的に想起できる顔の数と，潜在的に想起し認識できる顔の数の比を実験から求めることで，ヒトは平均して 5,000 人の顔を覚えることができると試算した．これらの研究から，顔画像は記憶しやすく，かつ大量に記憶できることがわかっている．

また，Willis と Todorov [44] によれば，人の顔画像は 100 ms の露光時間であっても，時間制約のない場合の印象と相関を示し，十分に印象を知覚できると報告している．また，Hancock ら [45] の研究によれば，既知の顔であれば不鮮明な顔画像であっても容易に同定できることを報告している．これらの研究から，顔画像は瞬時に知覚でき，既知の顔であれば容易に再認できることがわかっている．

2.3.4 顔画像のインジケーターへの応用

前項では顔画像に記憶と知覚の両面で利点があることを述べた．ここで，顔画像の記憶のしやすさを正確性，知覚のしやすさを利便性と捉えた上で，顔画像をインジケーターとして応用することを検討する．顔画像のインジケーターをもとに判断することは顔画像の正確性を活用することになり，安全性が見込まれる．また顔画像のインジケーターをユーザーが見るときには，顔画像の利便性が活用される．ここから，顔画像のインジケーターは顔画像の特徴を利用することで，情報セキュリティ分野ではトレードオフとなる安全性と利便性の双方を満たしている．つまり，顔画像をインジケーターに活用することは，認知心理学の観点から可能性のある選択肢である．

第 3 章

提案手法

前章では，フィッシング対策のための研究のうち，フィッシング攻撃に直面したユーザーの意思決定支援に着目した先行研究を概観した．また，通底する課題として一般的なユーザーには難解であることを指摘した．

本章では，NOPHICE（NOtify PHIshing attacks with faCial imagEs）システムを提案する．NOPHICE システムは，ユーザーが閲覧したウェブサイトの URL から，敵対的生成ネットワーク（Generative Adversarial Network, GAN）を利用してヒト顔画像を生成し，生成された顔画像をユーザーに提示する．本システムを利用してフィッシングサイトを判別するにあたって，ユーザーは普段利用するサービスの正規のウェブサイトに対応する顔画像を予め記憶していることを前提条件とする．そのうえで，ユーザーが閲覧しているウェブサイトが正規のウェブサイトかフィッシングサイトであるかを判断する際に，正規のウェブサイトと関連させて既にユーザー自身が記憶している画像と，システムから提示された画像とが，一致するかを判断材料とすることを狙いとしている．

3.1　GAN を利用したヒト顔画像生成

近年機械学習分野の発展は目覚ましいものであり，特にニューラルネットワーク及びディープニューラルネットワークによって諸分野への応用が進んでいる．中でも，互いに競合する 2 つのニューラルネットワークによって実装される GAN では，学習済み生成モデルにノイズを入力することで，学習に用いられたデータに近い疑似データを生成するものである [46]．フィッシングの検知に機械学習を応用した先行研究は数多くあり [47–49]，一部は GAN を用いている [50, 51]．

本システムでは，ヒトの顔画像の生成に GAN による学習済み生成モデルを用いる．GAN を用いる利点として 2 つの理由が挙げられる．第一に，生成モデルへの入力値であるノイズを変化させることで数に制限なく顔画像を生成できることが挙げられる．無数に存在するウェブサイトのそれぞれに対して，個別に対応した顔画像を必要に応じて生成して確保でき

図 9　StyleGAN2 で生成した顔画像例

ることに繋がる．第二に，実在する人の画像をそのまま使用せずに，生成された疑似データを利用できることが挙げられる．実在の人物の写真画像を利用すると，その画像と関連付けられた特定のウェブサイトの善性あるいは悪性に直接結びついてしまう形になり，社会的名声を傷つけてしまう事になりかねない．

例えば Karras らによる StyleGAN では，図 9 のように実際の顔画像と判別のつかない疑似顔画像を生成されることが報告されている．

3.2　システムの設計

本論文で提案するシステムは図 10 に示すような構造を持つものである．このシステムは次のような手順で動作する．

(1) ウェブブラウザで閲覧しているウェブサイトから URL，ドメイン名といったウェブサイトの同一性を示す情報を抽出し，ノイズを生成する．
(2) 生成されたノイズを GAN による学習済みモデルに入力し，ヒトの顔画像を生成する．
(3) 学習済みモデルはウェブサイトに対応付けられて生成されたヒトの顔画像をユーザーに提示する．
(4) ユーザーは提示された画像を元に，現在閲覧しているウェブサイトが正規のウェブサイトであるか，フィッシングサイトであるかを判断する．

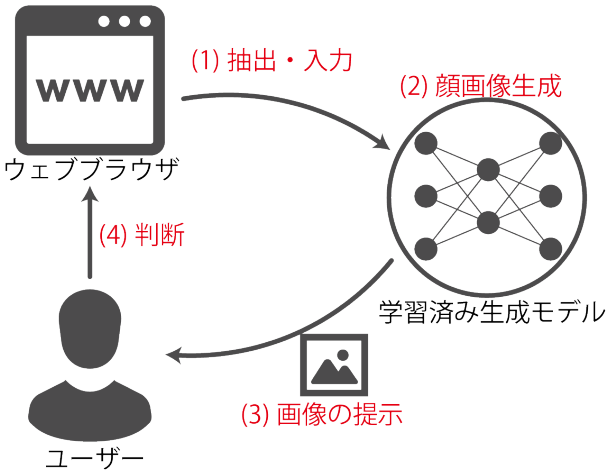


図 10　提案システムの設計図

3.3　システムの実装

提案したシステムを図 11 のように実装した．画像の生成にあたっては一定水準のリソースを要求されるため，ブラウザの拡張機能としてユーザーインターフェース部分を担当するインターフェース部分と，生成済みモデルによる顔画像生成を担当するモデル部分の，大きく 2 つに分解して提案したシステムを実装した．

このシステムは具体的に次のような手順で動作し，ユーザーに顔画像を提示する．

(1) ユーザー操作によって，現在閲覧しているウェブサイトの URL がインターフェース部分からモデル部分に送られる．
(2) モデル部分では，インターフェース部分から送られた URL を元に顔画像を生成し，インターフェース部分に顔画像を返す．
(3) インターフェース部分は，モデル部分から送られた顔画像をユーザーに提示する．

インターフェース部分とモデル部分の詳細な実装は以下に記す．

3.3.1　インターフェース部分

ブラウザの拡張機能としてユーザーインターフェースに関する機能を担当するインターフェース部分は，まずユーザーが拡張機能のエリアに表示されるボタンをクリックすることで一連の動作が始まる（図 12）．これにより，インターフェース部分はユーザーが現在閲覧しているウェブサイトの URL を取得しモデル部分に送る．その後，モデル部分から顔画像が返されると，図 13 のようにブラウザの拡張機能の領域上に顔画像を表示する．インターフェース部分の動作を擬似コードで示すとアルゴリズム 1 のようになる．

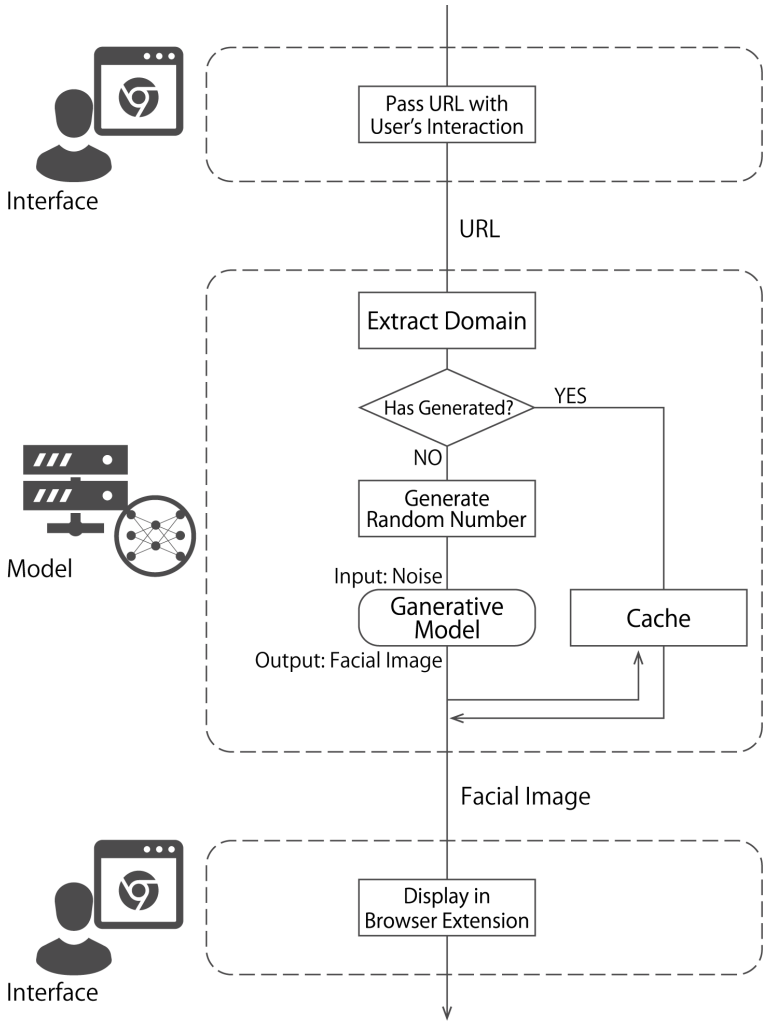


図 11　提案システムの実装図

3.3.2　モデル部分

顔画像を生成するモデル部分は次のように動作する．まず，インターフェース部分から送られた URL から Fully Qualified Domain Name (FQDN) を抽出する．顔画像はこの FQDN を元に生成される．FQDN を参照し対応した顔画像が生成済みでなければ，次の手順として FQDN から顔画像生成モデルに入力する乱数列を生成する．ここでは，可変長である FQDN を取り扱いやすいように SHA-256 関数にかけることで固定長とする．この FQDN のハッシュ値を乱数生成器のシード値として，疑似乱数列を生成する．そして，生成された乱数列をノイズとして，学習済み顔画像生成モデルに入力する．今回の実装では学習

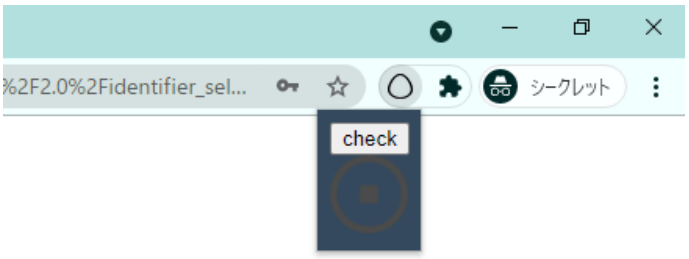


図 12　起動時の拡張機能の様子．check のボタンをクリックすることで次の手順に進む

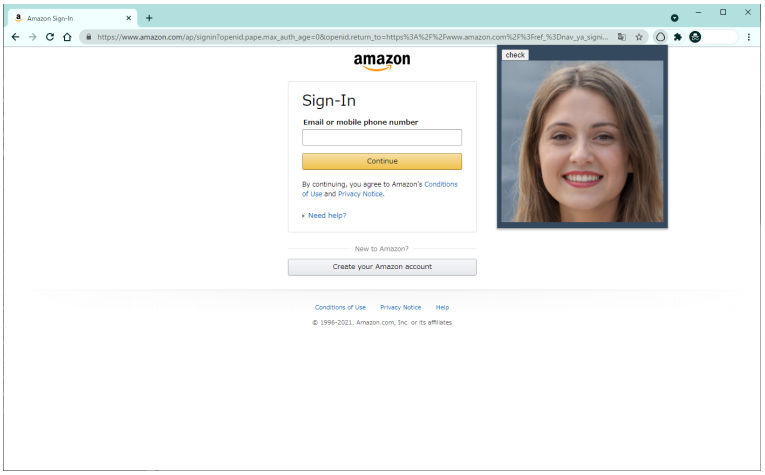


図 13　提案システムでの画像表示例．FQDN(www.amazon.com) に対応した画像はブラウザウィンドウの右上エリアに表示される

済みモデルに StyleGAN2 によって事前に学習されたものを用いた．[*1]出力された顔画像は FQDN と対応付けて保存し，最後に生成した顔画像をインターフェース部分に受け渡す．一方で，FQDN を参照し対応した顔画像が生成済みであれば，保存された生成済み顔画像をキャッシュのように利用しインターフェース部分に受け渡す．このモデル部分の動作を擬似コードで示すとアルゴリズム 2 のようになる．

このような実装により，FQDN を元に乱数を生成し顔画像の生成モデルに入力することによって，文字情報として類似したドメインであっても，全く異なる顔画像が生成されることになる．これによって，IDN ホモグラフィック攻撃を含むホモグラフィック攻撃に対して，全く違った画像をユーザーに提示することを可能としている．

[*1] 利用した学習済みモデルは Karras らによって利用可能となっている．`https://nvlabs-fi-cdn.nvidia.com/stylegan2/networks/stylegan2-ffhq-config-f.pkl`（2021/08/20 現在）

アルゴリズム 1 Interface Part

```
if check_button.clicked then
    URL ← ExtructUrl(browser.currentPage)
    send URL to ModelPart
    recieve face_image from ModelPart
    display face_image
end if
```

アルゴリズム 2 Model Part

```
recieve URL from InterfacePart
if image.associated_with(URL.domain) exsits then
    face_image ← cached_image[URL.domain]
else
    noise ← GenerateRandomNumbers(URL.domain)
    face_image ← GenerativeModel(noise)
end if
send face_image to InterfacePart

function GenerateRandomeNumbers(domain)
    hashed_domain ← SHA256(domain)
    for i = 0, 1, 2, 3 do
        seeds[i] ← hashed_domain[32i : 32i + 31].to_int
    end for
    for i = 0, 1, . . . , NoiseSize − 1 do
        random_numbers[i] ← Random(seeds)
    end for
    return random_numbers
end function
```

3.3.3 提案システムの検証環境

提案システム NOPHICE が動作するか検証した際の動作環境を次に示す.

- GPU: NVIDIA GeForce GTX 1070 Ti
- OS: Windows10 20H2

- ブラウザ: Windows 版 Google Chrome 87
- Python3.7.8/CUDA10.0/cnDNN7.5/Tensorflow1.14

この動作環境において，Alexa の TOP50 [52] のドメインに対して NOPHICE で顔画像を生成したところ，一つの顔画像生成にかかる時間は平均 24.34 秒（23.75 s–26.05 s）であった．なお，既にドメインに対応する画像が生成済みであった場合，先に述べた生成済み画像の流用によって画像生成に要する時間を短縮することが可能である．

第 4 章

調査

本章では，前章で提案した NOPHICE システムがフィッシングの判別に有効であるか検証するために実施した，調査の内容とその結果について述べる．

4.1 調査の目的

調査では以下に挙げる仮説を検討することにより，提案手法の有効性をセキュリティと利便性の 2 つの観点で検証することを目的とする．

帰無仮説 H_0^1　提案手法はユーザーが正しく判別できる割合を変化させない．
帰無仮説 H_0^2　提案手法はユーザーが誤って判別してしまう割合を変化させない．
帰無仮説 H_0^3　提案手法はユーザーが判別するときの主観評価を変化させない．

これらの帰無仮説に対応する対立仮説は次のとおりである．

対立仮説 H_1^1　提案手法はユーザーが正しく判別できる割合を変化させる．
対立仮説 H_1^2　提案手法はユーザーが誤って判別してしまう割合を変化させる．
対立仮説 H_1^3　提案手法はユーザーが判別するときの主観評価を変化させる．

以降では，正規のウェブサイトを正規のウェブサイトであるとユーザーが判定するか，フィッシングサイトをフィッシングサイトであるとユーザーが判定し，ウェブサイトを正しく判別した割合を正答率と定義する．また，正規のウェブサイトをフィッシングサイトであるとユーザーが判定するか，フィッシングサイトを正規のウェブサイトであるとユーザーが判定し，ウェブサイトを誤って判別してしまった割合を誤答率と定義する．

4.2 調査の概要

本研究では，株式会社マクロミルを通じ，16 歳〜59 歳の男女合わせて 110 名に協力のもと，2021 年 3 月にウェブアンケート調査を実施した．なお，調査を実施するにあたって東

表 1　実験参加者の属性区分における人数

属性		A (提案群)	B (対照群)
年齢区分	16 – 19	11	11
	20 – 29	11	11
	30 – 39	11	11
	40 – 49	11	11
	50 – 59	11	11
性別	男性	22	18
	女性	33	37
合計		55	55

京大学ライフサイエンス研究倫理支援室を通じて研究倫理審査申請を行っており，承認を得ている．また，調査を依頼した企業では親権者の同意の上で，未成年者による回答が行われており，アンケート調査においても個人を特定可能な情報を収集せず，すべての年代に対して適切に実施できるように配慮した．18 歳以下のユーザーであっても，フィッシング攻撃の対象となり個人情報を窃盗される可能性があるため，アンケート調査の対象から排除しなかった．

4.3　調査方法

参加者を 55 名ずつの二群に分け，それぞれのグループに対してアンケートへの回答を要請した．参加者の年齢層ごと並びに性別ごとの人数は表 1 に示すとおりである．アンケートの回答時間について制限を設けず，PC やスマートフォンといった回答に利用するデバイスについても指定しなかった．また，アンケートに回答する参加者個人の状況についても指定しなかった．

提案システムの実装を用いたグループをグループ A とし，通常のブラウザを用いたグループをグループ B とする．

アンケートでは，ウェブサイトを表示したブラウザのスクリーンショット画像を見て参加者にフィッシングサイトかどうか判別してもらった．画像はすべて横 1920px 縦 1050px の PNG 画像である．使用したブラウザは Windows 版 Google Chrome であり，ブックマークバーは表示せず，判別するウェブサイトのみを 1 つのタブで開いている画像である．

また，フィッシングサイトの画像は，実際のフィッシング攻撃に利用されていたフィッシ

ングサイトにアクセスして取得したスクリーンショットである．実在のフィッシングサイトを実験に用いることで，実際にフィッシング攻撃に直面したユーザーの判別精度を再現しようとするものである．また，ウェブサイトではなくスクリーンショット画像を利用することで，実験参加者が誤って個人情報を入力してしまう可能性を排除した．

4.4 アンケートの質問内容

グループ A に対しては提案システムの実装を用いた画面を，グループ B に対しては提案システムを用いず通常のブラウザの画面を使用した．グループ A の参加者のみに対して，アンケートの最初に提案システムの説明を図 13 とともに行った．アンケートは，(1) 正規のウェブサイトの記憶，(2) ウェブサイトの判別のテスト，(3) ユーザーによる評価の 3 つで構成される．なお，アンケートの冒頭では，本アンケートで個人を特定できる情報を収集せず，情報セキュリティの研究のみに利用されることを参加者に説明した．参加者は，アンケートの実施に同意した上でアンケートに回答した．

4.4.1 正規のウェブサイトの記憶

正規のウェブサイトのログイン画面を表示したブラウザの画像を参加者に 4 枚提示し（図 14），それぞれの画像を記憶するように指示をした．

4.4.2 ウェブサイトの判別のテスト

ウェブサイトのログイン画面を表示したブラウザの画像を参加者に提示し，その画像が正規サイトのものであるか，フィッシングサイトのものであるかを判別させた．このテストは各アンケートにつき 10 問ある．各問題で提示される画像は，正規のウェブサイトを表示した (1) と同一の画像，もしくは実際のフィッシングサイトを Web ブラウザに表示した画像である．判別に用いたフィッシングサイトは，(1) で参加者に提示した 4 つの正規ウェブサイトのいずれかのブランドを騙るものに限られている．

4.4.3 ユーザーによる評価

(1)，(2) を通じて参加者の主観評価を調査するために表 2 に示す質問を行った．質問 1 では判別の難易度を，質問 2 では正規のウェブサイトの記憶にあたっての心理的負担を，質問 3 では判別した際の自信についての評価を，質問 4 では判別にあたっての心理的負担を問うた．これらの質問では，各記述にどの程度同意できるかを 5 段階のリッカート尺度で質問した．

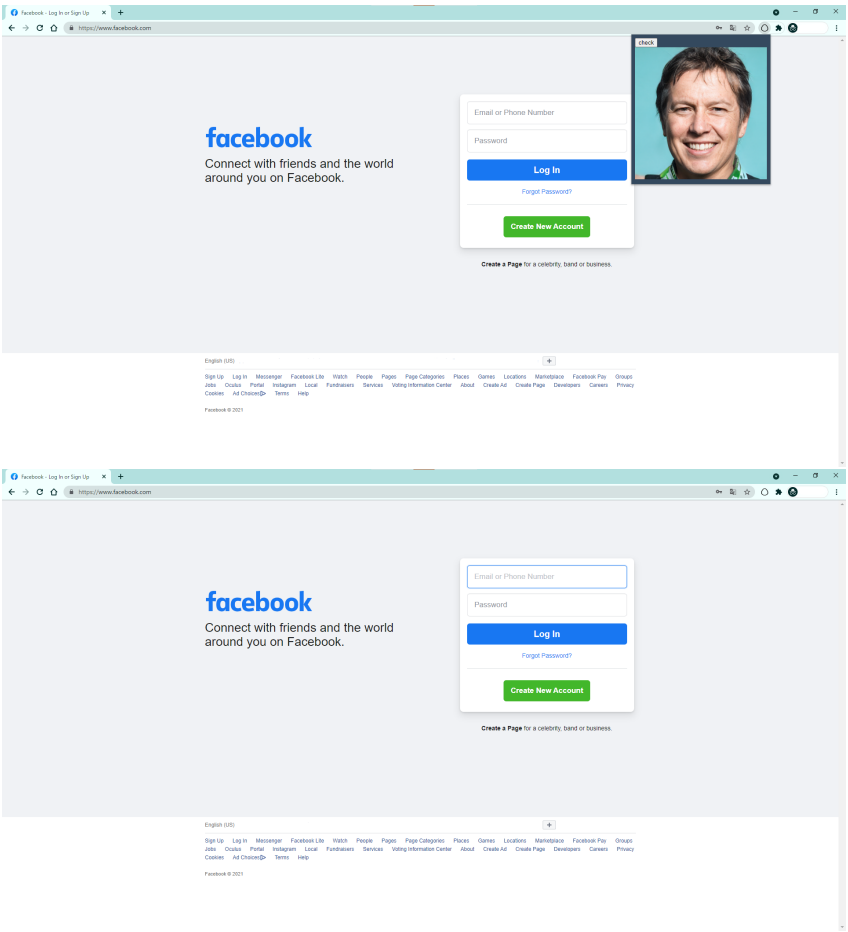


図 14　実験参加者に正規ウェブサイトとして提示した画像の例．これらは同じウェブサイトであり，グループ A には上図をグループ B には下図を提示した

表 2　ユーザー評価におけるアンケート項目

ユーザー評価の質問項目
Q1. 判別は容易であった.
Q2. 覚えやすかった.
Q3. 自身を持って判別できた.
Q4. 判別に時間や手間はかからなかった.

5 段階リッカート尺度: (1) 全くそう思わない – (5) 強くそう思う

表 3　正答率及び誤答率の両側 t 検定の結果（*: $p < 0.05$, **: $p < 0.01$）

種別		提案群平均値（SD）	対照群平均値（SD）	p
正答率	正規	0.518(0.396)	0.427(0.371)	0.217
	フィッシング	0.636(0.425)	0.555(0.391)	0.296
	全体	0.589(0.356)	0.504(0.307)	0.181
誤答率	正規	0.159(0.277)	0.282(0.326)	0.036*
	フィッシング	0.052(0.139)	0.188(0.266)	0.001**
	全体	0.095(0.141)	0.225(0.204)	< 0.001**

4.5　調査結果

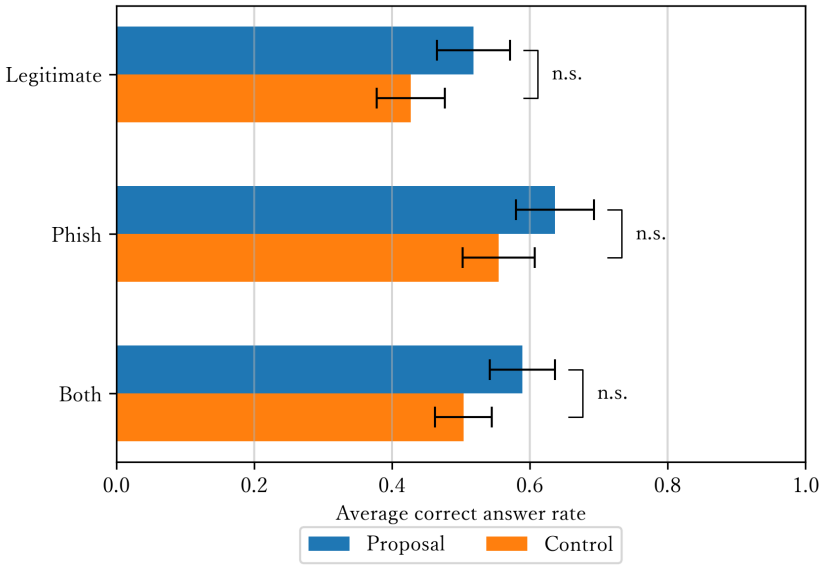
提案法を利用したグループ A（以下，提案群）と，既存の方法を利用したグループ B（以下，対照群）のそれぞれから 55 名ずつの回答が得られた．なお，本節で参照する図中のエラーバーは標準誤差を示す．

4.5.1　正答率の比較

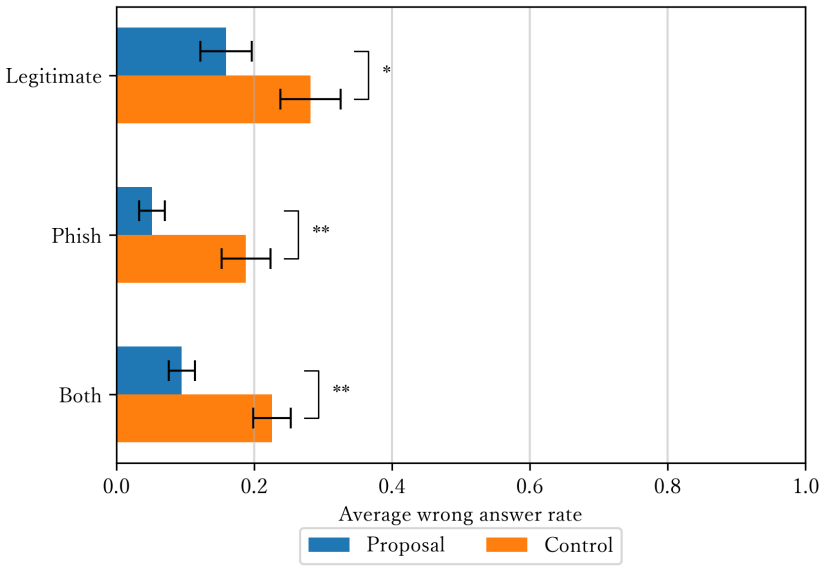
正答率の平均値について，正規ウェブサイト，フィッシングサイト及びその両方での，提案群と対照群との比較を図 15a に示す．また，正答率について両側 t 検定を行った結果を表 3 に示す．正答率の平均値について，いずれの場合も提案群は対照群よりも大きい値を示した．しかし，t 検定の結果得られた p 値は有意水準 $\alpha = 0.05$ よりも大きく，統計的有意差は認められなかった．従って，アンケート調査の結果からは帰無仮説 H_0^1 を棄却できず，対立仮説 H_1^1 が正しいと結論づけられなかった．

4.5.2　誤答率の比較

誤答率の平均値について，正規ウェブサイト，フィッシングサイト及びその両方での，提案群と対照群との比較を図 15a に示す．また，誤答率について両側 t 検定を行った結果を表 3 に示す．誤答率の平均値について，いずれの場合も提案群は対照群よりも小さい値を示した．また，t 検定の結果得られた p 値は有意水準 $\alpha = 0.05$ よりも小さく，統計的有意差が認められた．従って，アンケート調査の結果からは帰無仮説 H_0^2 が棄却され，対立仮説 H_1^2 が正しいと結論づけられた．また，提案群の誤答率の平均値は対照群のそれよりも小さい値を示したため，提案手法はユーザーが誤って判別してしまう割合を減少させると結論づけられる．



a: 正規・フィッシング・両方の正答率．有意差はなかった．



b: 正規・フィッシング・両方の誤答率．提案群は対照群に比べて全ての項目で優れたスコアとなった．

図 15　提案群と対照群の比較．（n.s.: 有意差なし, *: $p < 0.05$, **: $p < 0.01$）

表 4　ユーザー主観評価の両側 t 検定の結果

質問項目	提案群平均値（SD）	対照群平均値（SD）	p
Q1	2.455(1.033)	2.200(1.025)	0.197
Q2	2.436(0.976)	2.327(0.943)	0.552
Q3	2.291(0.993)	2.236(1.137)	0.789
Q4	2.655(1.040)	2.527(1.288)	0.570

5 段階リッカート尺度: (1) 全くそう思わない – (5) 強くそう思う

4.5.3　ユーザーの主観評価

提案群と対照群のユーザーによる主観評価の比較を図 15 に示す．また，ユーザー評価について両側 t 検定を行った結果を表 4 に示す．なお，質問内容は表 2 に示したとおりである．各質問の平均スコアは，リッカート尺度の選択肢に 5（強くそう思う）から 1（全く思わない）までのスコアを割り振った上で質問ごとのスコアの平均値を算出した値であり，スコアが大きいほど，質問項目に対して同意できる傾向が強いことを表している．各設問について提案群が対照群よりもより同意できる値を示した．しかし，t 検定の結果得られた p 値は有意水準 $\alpha = 0.05$ よりも大きく，統計的有意差は認められなかった．従って，アンケート調査の結果からは帰無仮説 H_0^3 を棄却できず，対立仮説 H_1^3 が正しいと結論づけられなかった．

4.6　まとめ

本章では，提案システム NOPHICE の有効性を検証するために，調査を実施し結果を分析した．

調査では NOPHICE の利用によって，フィッシングサイトを正規のウェブサイトと判断してしまう誤答率が有意に減少した．また，フィッシングサイトや正規のウェブサイトを正しく判別できる正答率は，有意差は無かったものの増加した．フィッシングサイトを正規のウェブサイトと誤認して個人情報を入力してしまうという，ユーザーがフィッシングの被害を受けるシナリオを考えると，正答率が向上しなくとも，誤答率が減少すればフィッシングの被害抑制に繋がると考えられる．よって，NOPHICE の利用によってユーザーの安全性は向上した．一方で，利便性を問うたユーザーの主観評価では有意差は無かったものの，NOPHICE の利用によって同意できる傾向が強くなった．よって，少なくとも NOPHICE の利用によってユーザーの利便性は損なわれなかった．

調査の結果をまとめると，利便性と安全性のトレードオフが一般的な情報セキュリティ

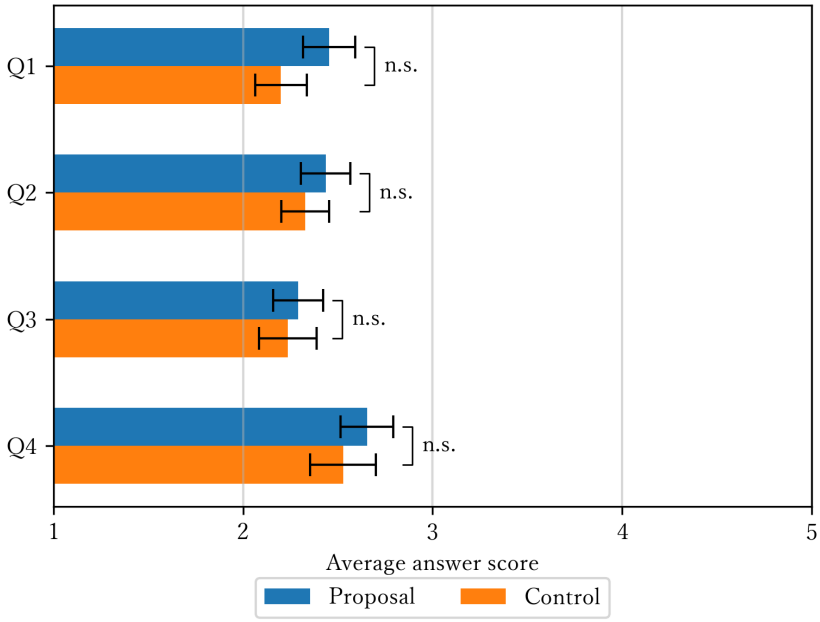


図 16　ユーザー評価の提案群と対照群の比較（n.s.: 有意差なし）

において，利便性を損なわず安全性を向上させた NOPHICE は有効であったと結論づけられる．

第 5 章

議論

本研究では，フィッシング対策のためのユーザー支援手法として顔画像を利用することを提案した．そして提案事項を NOPHICE システムとして実装し，ユーザー調査を行った．提案手法を発展させるため，本章では，現在の NOPHICE の実装を基に応用や問題点，課題など多角的な視点で議論する．

5.1 提案手法の特徴と応用

5.1.1 画像を用いる事によるアクセシビリティ

本研究では，フィッシングサイトにアクセスしたユーザーが，普段利用しているウェブサイトと違うことを，特別な知識なしに感知できるような新しい形式のインジケーターを模索した．そして一つの結論として，本研究では顔画像を利用するシステムを提案した．

ここで特別な知識とは，従来フィッシングサイトの判別基準に用いられてきた URL，ドメイン名，SSL 証明書などの情報 [28] に対する知識である．こうした従来の判断基準となる情報は文字列の形式に依存している．文字列によってフィッシングサイトかどうか判断することは，読み書きの能力に著しい困難を持つディスレクシア（失読症，難読症）のユーザーにとって難しい課題であると考えられる．アメリカでは 5% 弱の児童がディスレクシアによって特別な配慮がされた教育を受けており [53]，15–20% の児童はディスレクシアの症状があると予想されている [54]．また，日本においても 8% 程度がディスレクシアの症状を有していると報告されている [55]．よりよい情報セキュリティを考えるにあたって，ディスレクシアのユーザーを見過ごしてはならず，顔画像を用いた提案システムは一つの回答たりうる．

一方で，相貌失認という障害を抱えるユーザーには，提案システムが有効に機能しない可能性がある．脳の器質的障害が見られないにもかかわらず顔の認識能力が著しく低い症状を，先天性相貌失認（発達性相貌失認）と呼び，2–2.5% 程度の割合で存在すると推定されている [56]．相貌失認を抱えるユーザーは，提示された顔画像を見分けることができない可能性があるため，むしろ URL や SSL 証明書表示といった文字列情報の方が，フィッシング

サイトの判別に利用しやすいと推測される．

5.1.2 生成された顔画像の共有

提案システムの実装では，特定のドメイン名に対応する顔画像は全てのユーザーであるため，生成済み顔画像のキャッシュを共有することで，画像表示までの時間を短縮するような応用が可能である．一方で，他のユーザーが一度利用したウェブサイトのキャッシュも機能するため，少数の人数が提案システムを利用していたときに，ユーザーが提案システムを利用した際の画像表示までの時間から，別のユーザーがそのウェブサイトを利用していたという閲覧履歴が間接的に漏洩する可能性が想定される．

5.1.3 FQDN ではなくドメイン名を元に画像生成する手法について

提案システムの実装では FQDN を元に顔画像を生成しているが，サブドメイン名を含めずにドメイン名を元に顔画像を生成する手法が考えられる．長所としてはロードバランサーや複数サービスを展開するためにサブドメイン名が変化する場合であっても，システムはブランドに紐付いた共通の顔画像を提示することができ，ユーザーにとって識別が容易になることが考えられる．一方で，短所としてクラウドサービス上にフィッシングサイトを設置されていた際のシステムの利用が考えられる．仮に，正規ウェブサイト `legit.cloud.example.com` とフィッシングサイト `phish.cloud.example.com` に対して，ドメイン名 `cloud.example.com` を元に顔画像の生成がなされると，双方が同一の顔画像となってしまい判別できなくなってしまう．このようなクラウドサービスを利用したフィッシングの調査は本研究の範囲外となるが，本研究の実装では FQDN を元に画像生成を行った．

5.1.4 検知技術との組み合わせによる応用

閲覧しているウェブサイトがフィッシングサイトであるかどうかを判別する既存の技術と組み合わせ，フィッシングサイトである確率といった怪しさに関する数値を視覚的な情報としてユーザーに提示できるか，応用を検討する．生成される顔画像が示す属性，例えば年齢や性別をフィッシングサイトであるかどうかに対応付けて操作する方法が考えられる．しかし，人の属性とフィッシング行為の善悪を対応付けることに繋がるため，好ましくない．同様に，色相や明度を対応付けることも好ましくない．ここで，彩度を対応付けることを考え


図 17　生成された顔画像（左図）とフィッシングの怪しさの度合いに合わせて彩度を調整した画像（右図）

る．例えば，フィッシングサイトであるか確からしさを判別する既存の技術[*2]と組み合わせ，その値によって生成された顔画像の彩度を減少させる．図 17 は怪しいウェブサイトであるとして彩度が大きく下げられた画像の調整例である．他にも，フィッシングサイトである確からしさに応じてぼかし等の画像効果を加える方法が考えられる．

他方で，Allow List や Deny List といった技術との組み合わせを考えた場合，それらの検知技術に許可/拒否されたウェブサイトについては，特定の顔画像を統一して表示することが考えられる．

5.2　提案手法の限界

5.2.1　ユーザーごとのばらつき

アンケート調査の回答結果を分析すると，提案法を用いても全く判別できず，全ての設問に対してわからないと回答したユーザーが存在した．一方で既存の方法においても，十分な精度で適切に判別できるユーザーも存在する．このようにユーザーの能力にばらつきがあるため，画一的な方法では限界が生じてしまう．

フィッシング攻撃がソーシャルエンジニアリングを用いる攻撃である以上，いずれかのタイミングでユーザーによる判断が生じる．そのため，能力の異なるユーザーそれぞれがより安全な判断できるように，フィッシング攻撃の判別を支援する方式をシステムが複数提供で

[*2] 今回の実装では GitHub 上に公開されている実装を利用した．`https://github.com/zpettry/AI-Deep-Learning-for-Phishing-URL-Detection`（2022/01/15 現在）

きる仕組みを作っていくことが望ましい．そして，システムから提供された複数のユーザー支援手法から，ユーザーが自身に即した手法を選ぶことで，よりよい体験に繋がると考えられる．

5.2.2 ユーザーにとって未知のサービスを騙った攻撃

提案システムは，ユーザーが普段利用しているサービスのウェブサイトと，そのサービスを騙った偽造ウェブサイトの判別を支援することを狙いとしている．つまり，ユーザーにとって既知のサービスを騙った攻撃に対して有効である．特定のサービスの ID やパスワードを窃取するには，前提として攻撃対象が既にそのサービスを利用しアカウントを作成している必要がある．ユーザーが既にサービスを利用しており，その既知のサービスを騙ったフィッシングサイトに個人情報を入力してしまう流れが典型的である．

翻って，ユーザーにとって未知のサービスを騙った攻撃について検討したい．未知のサービスであれば，本来ユーザーが入力すべき情報が存在しないはずである．そのため一般的には攻撃が成立しにくいと考えられる．しかし，銀行の口座情報やクレジットカード番号などの複数のサービスに共通して入力する状況が存在する個人情報は，未知のサービスであっても入力してしまう可能性がある．また，ID やパスワードを複数のサービスで使いまわしているユーザーが，未知のサービスにも関わらず認証情報を入力してしまう可能性がある．

提案システムでは，こうした未知のサービスを騙った攻撃に対して有効に機能しない可能性がある．提示された画像の比較元となる正規のウェブサイトの顔画像をユーザーが知らず，そもそも比較が成り立たないからである．しかし，未知のサービスであれば正規のウェブサイトでもフィッシングサイトであっても，ユーザーにとって未知の顔画像をシステムが提示することになる．ユーザーが記憶している正規のウェブサイトの顔画像群に含まれないことから，ユーザーがフィッシングサイトであると判断することは可能である．

もしまだ検知されていないフィッシングサイトが攻撃に利用されていたならば，ユーザーが現在のウェブブラウザの実装でフィッシングサイトを判別することは，表示されているサービスがユーザーにとって既知であるか未知であるかに関わらず，難しい課題である．なぜならば，フィッシングサイトがまだ検知されていない状況では，現在のウェブブラウザがフィッシング攻撃を警告する手段を持ち合わせていないからである．ユーザーにとって未知のサービスを騙ったフィッシング攻撃に対する警告の仕方は，既存のウェブブラウザも提案システムも同様に課題である．

5.2.3 ユーザーの利用するデバイスの画面サイズ

本研究の調査では，横 1920 px 縦 1280 px の解像度の画面で PC 版ウェブブラウザをウインドウ最大化させた上で，提案システムを起動した様子のスクリーンショットを使用した．画面の大きさと顔画像の提示方法についていくつか懸念事項がある．

第一に，4K に代表されるような高解像度の画面において顔画像が小さく表示されてしまい，判別が難しくなる可能性がある．この問題は画面解像度に連動させて，顔画像の表示をスケールすることで対応できる．第二に，スマートフォンのように物理的な画面サイズが小さい端末で，顔画像を表示する場合を考えたい．現在の NOPHICE システムの実装のように右上の領域に顔画像を提示しようとすれば，表示しているウェブサイトのコンテンツに意図せず被ってしまう，あるいは被ることを避けようとして顔画像が小さくなってしまうことで，ユーザーにとって不便になる．画面サイズが限られており，顔画像のサイズを確保すればウェブサイトのコンテンツと被ることは避けられないため，むしろスマートフォンで広く用いられる警告画面と同様の全画面のインターフェースを採用する方が，他のユーザーインターフェースと一貫性を持ち，ユーザーの利便性を損なわない可能性がある．例えば，図 18 のような画面中央に顔画像を表示する実装が考えられる．このように，ユーザーの使用するデバイスによって適切な表示方法を検討することで，提案システムはより効果的になることが考えられる．

5.3 提案システムに対する攻撃について

5.3.1 類似画像の提示による欺瞞

提案システムに対する攻撃として，攻撃者側が同一もしくは類似の顔画像を提示することでユーザーに正規のウェブサイトであると誤認させることが考えられる．この類似の画像を提示する攻撃には，主に 3 つの方法が考えられる．（1）正規ウェブサイトから生成される画像と同一の画像を生成できるドメイン名でフィッシング攻撃を展開する（2）正規ウェブサイトから生成される画像と類似した画像を生成できるドメイン名でフィッシング攻撃を展開する（3）正規ウェブサイトから生成された画像をフィッシングサイト上に表示する

まず，正規のウェブサイトと同一の画像を生成するドメイン名の利用による攻撃を考える．この攻撃を成立させるには，正規のウェブサイトのドメイン名をキーとした時の SHA-256 ハッシュ値と同じ，SHA-256 ハッシュ値を出力するドメイン名を見つけ出せばよい．このことは，SHA-256 に対する第二原像攻撃に相当する．SHA-256 では弱衝突耐性の脆弱性は発見されていないため，総当たりで探索しなければならない．SHA-256 は 256bit のハッシュ値を出力するため，全通りのハッシュ値を探索するためには 2^{256} 回の計算が必要となる．

図 18　スマートフォンへの提案システムの実装イメージ例

ここで，ドメイン名には区切り文字であるピリオド（.）の他，大文字小文字を区別しないアルファベット（A–Z），数字（1–9），ハイフン（-）の計 38 種の文字が使える．ピリオドで区切られた部分はラベルと呼ばれ，ラベルはひとつあたり 63 文字以下でなければならないこと，ラベルの最初と最後にハイフンは使用できないことの 2 つの制約がある．簡単のため，フィッシング攻撃用ドメイン名の探索では，アルファベットと数字の 36 種の文字の組み合わせで行うとする．2^{256} 通りの探索に必要な文字列の長さ l は，

$$
\begin{aligned}
&36^l > 2^{256} \\
&\therefore l > \frac{256 \log 2}{\log 36} = 49.51\ldots
\end{aligned} \tag{1}
$$

より 50 文字とわかる．これはドメイン名の仕様に反しない形で実現可能な長さである．

一方で，2^{256} 回のハッシュ計算に要する時間を推算する．ハッシュ計算に特化した ASIC のうち，2022 年 1 月時点で一般に発表されている中で最速の製品である，Antminer S19 XP（Bitmain 社製）は 140 Th/s の速度で SHA-256 のハッシュ計算を行える [57]．この機器を 1 台用いて計算を行えば 8.27×10^{62} s $\approx 2.62 \times 10^{55}$ 年で探索が完了する．そのため，正規のウェブサイトと同一の画像を生成するドメイン名を用いたフィッシング攻撃は，現実的には不可能であると結論付けられる．

また，違うハッシュ値を出力するドメイン名であっても，生成される乱数列が正規のウェブサイトから生成される乱数列と同一であれば，顔画像生成モデルから同一の顔画像が生成される．本研究の実装では乱数生成器にメルセンヌツイスターの実装である MT19937 を用いており，擬似乱数列の周期は $2^{19937} - 1$ である．これは，入力するシード値の空間，すなわち SHA-256 のハッシュ値の空間 2^{256} に比べて非常に大きい．そのため，同じ乱数列を出力するドメイン名を見つけ出すことは難しく，攻撃方法として現実的ではないと結論付けられる．

次に，正規のウェブサイトと類似する画像を生成するドメイン名の利用による攻撃を検討する．攻撃者がいくつかのドメイン名で顔画像生成を試み，その中で類似する顔画像を生成したドメイン名でフィッシング攻撃を展開することを想定している．仮に，顔画像が表 5 のような成分要素の組み合わせで分類できるとする．ただしこれは筆者が試算のために大雑把に設定した値である．この表に従えば，顔画像は全部で $2 \times 4 \times 4 \times 8 \times 10 = 2560$ 通りに分類できる．これらの成分要素がすべて一致したときに類似の顔画像であるとすれば，現実的な試行回数で類似の顔画像を生成できる可能性がある．また，人は 5000 通りの顔を覚えられると試算した Jenkins ら [43] の研究からも，生成できる顔画像の数 2^{256} と比べて少ない，現実的な試行回数で類似の顔画像を生成できると予想できる．

最後に，正規のウェブサイトから生成される顔画像をフィッシングサイト上に表示することでユーザーを欺瞞する攻撃方法を検討する．提案システムの実装では，ウェブブラウザの拡張機能の領域上で画像を提示している．ウェブブラウザの実装に特別の脆弱性がなければ，フィッシングサイトから拡張機能の領域を操作することはできない．そのため，ウェブサイトの画面領域で画像を表示することで擬似的に提案システムを再現する方針が考えられる．この攻撃は，提示された顔画像が提案システムに依るものではないと気づけない，ユーザーの誤認次第で十分成立しうる．

表 5　顔画像の成分例

要素	分類数	取る値の例
性別	2	男性，女性
人種	4	ネグロイド，コーカソイド，モンゴロイド
年齢	4	子供，青年，壮年，老年
髪のタイプ	8	黒の長髪，白髪，金の短髪
顔のタイプ	10	細いあご，箱型

注：各値は試算のために筆者が設定した値である．

以上のように，類似の顔画像を提示する提案システムへの攻撃を検討し，一部の攻撃方法は成立する可能性があることがわかった．本研究の実装では，あるウェブサイトに対して提案システムがすべてのユーザーに共通の顔画像を生成する．そのために，ユーザーがどのような画像を元にウェブサイトの判別をするのか，攻撃者は事前に知ることができ，類似もしくは同一の顔画像を表示することによる攻撃が可能となっている．従って，こうした攻撃方法への対策として，ユーザーごとに異なった顔画像を生成する方法が考えられる．例えば，ドメイン名の前にユーザーごとに固有な文字列を追加してから，顔画像生成モデルに入力する．このような方法によって，攻撃者が類似もしくは同一の顔画像を提示することを非常に難しくすることができる．

ただし，類似の画像の提示によってユーザーを欺瞞し提案システムを攻撃することに成功したとしても，正規のウェブサイトに類似したドメインを利用する従来の攻撃手法の要件と両立させることは，攻撃者にとって更に難しい障壁となる．つまり，類似の顔画像の提示と類似のドメインの利用を両立したフィッシング攻撃は，現実的に成立する攻撃にはならないだろうと考えられる．

5.4　普及における課題

本節では，仮に提案システムを普及させるとした時の課題や障害を，いくつかの立場から考察する．

5.4.1　ウェブサービス提供者の負担

提案システムはウェブサービス提供者に協力を要請することはないため，特筆すべき課題点はない．例えば，ドメイン登録時や SSL 証明書発行時に併せて顔画像を生成し登録する

ことや，ユーザーが接続した時にレスポンスに画像を含めることで，ウェブサービス提供者が提案システムへ関与する方針がある．しかし，登録や管理のコスト，画像を送信することによるトラフィックの圧迫，フィッシング攻撃者が正規のウェブサービスからユーザーに送られる顔画像を不正に取得することによるなりすまし，といった問題が考えられる．こうした欠点から，提案システムの普及にあたってウェブサービス提供者は追加の作業を必要としないと，本研究は想定している．

5.4.2 ユーザーの負担

本研究では，提案システムの処理はユーザー側に大きく負担してもらうことを想定している．2022 年現在では，画像の生成にコンピューターの処理能力を要し，一般に用いられているノート型パソコンでは提案システムの利用に性能が不足している．ただ，ムーアの法則に表されるような半導体技術の進歩によって，性能不足の問題は解消されると予想できる．一方で，スマートフォンやタブレットなどのモバイル端末による提案システムの利用を考えると，発熱や消費電力の観点から，端末内での画像生成は避けたい．そこで，エッジコンピューティングを利用する方法が考えられる．提案システムのモデル部分の処理をエッジサーバーで行うことで，モバイル端末の負担を削減できる．例えば，第 5 世代移動通信システム（5G）に接続する端末が，5G 上で提供されるエッジコンピューティングを利用して提案システムを利用する．5G を提供する事業者やそのシステムに負担がかかってしまうものの，トラフィックの圧迫は最小限に抑えることができる．

第 6 章

結論

本研究では，偽造されたウェブサイトの判別にあたってのユーザーの意思決定を支援する既存の手法が難解であるとして，ヒトの顔画像を用いてユーザーの意思決定を支援するシステム NOPHICE を提案した．提案システムはモデル部とインターフェース部に分けて実装した．モデル部では，GAN による学習済みモデルを利用し，ウェブサイトのドメイン名に対応する顔画像を生成した．またインターフェース部は，ウェブブラウザの拡張機能として実装した．

そして提案システムの効果を検証するために，アンケート調査会社を通じて 110 名に協力いただき調査を実施した．調査の結果，提案システムによってウェブサイトを正しく判別できる割合が向上し，特にフィッシングサイトを正規のウェブサイトと誤認する割合は有意に減少した．また，利便性を問うたユーザーの主観評価も提案システムの利用によって向上した．提案システムによってユーザーの利便性と安全性が向上したことから，提案システムの有効性が示された．

さらに，提案システムのアクセシビリティ，検知システムとの組み合わせによる提案システムの応用，提案システムの限界，提案システムに対する攻撃，提案システムを普及させるにあたっての課題といった事項について，提案システムを広範な視点で議論した．スマートフォンのようなモバイルデバイス向けに提案システムを応用することや，ユーザーに能力のばらつきがあることからシステムは多様な選択肢を提示することは，議論の結果から導かれた今後の課題である．

本研究は，一般のユーザーにも活用できる水準であることに焦点を当ててシステムを設計した，初めてのフィッシング研究である．情報技術に精通している研究者は，しばしば一般的なユーザーの情報技術に対する理解度を見誤ってしまう．今後のセキュリティを考えるにあたって私達は，ユーザーの実態に即した方法を模索しなければならない．

参考文献

[1] James Nicholson, Lynne Coventry, and Pam Briggs. Can We Fight Social Engineering Attacks By Social Means? Assessing Social Salience as a Means to Improve Phish Detection. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pp. 285–298, Santa Clara, CA, July 2017. USENIX Association.

[2] Melanie Volkamer, Karen Renaud, Benjamin Berens, and Alexandra Kunz. User Experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn. *Computers & Security*, February 2017.

[3] Justin Petelka, Yixin Zou, and Florian Schaub. Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, New York, NY, USA, May 2019. Association for Computing Machinery.

[4] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. Your Attention Please: Designing Security-Decision UIs to Make Genuine Risks Harder to Ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, New York, NY, USA, July 2013. Association for Computing Machinery.

[5] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving SSL Warnings: Comprehension and Adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 2893–2902, New York, NY, USA, April 2015. Association for Computing Machinery.

[6] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Emre Acer, Elisabeth Morant, and Sunny Consolvo. Rethinking Connection Security Indicators. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*, SOUPS '16, pp. 1–13, USA, June 2016. USENIX Association.

[7] Timothy F. Brady, Talia Konkle, George A. Alvarez, and Aude Oliva. Visual Long-term Memory Has a Massive Storage Capacity for Object Details. *Proceedings of the National Academy of Sciences*, Vol. 105, No. 38, pp. 14325–14329, September 2008.

[8] コインチェック株式会社. 仮想通貨 NEM の不正送金に関するご報告と対応について.

https://corporate.coincheck.com/press/zlqckpSm, March 2018. (Accessed on 01/22/2022).

[9] Apple(日本). 著名人の写真に関する調査状況の報告. https://www.apple.com/jp/newsroom/2014/09/02Apple-Media-Advisory/, September 2014. (Accessed on 01/22/2022).

[10] BBC ニュース. サイバー被害の米パイプライン、身代金 4.8 億円の支払い認める 米紙報道. https://www.bbc.com/japanese/57181463, May 2021. (Accessed on 01/22/2022).

[11] 廣瀬陽子. ハイブリッド戦争 ロシアの新しい国家戦略 (講談社現代新書). 講談社, February 2021.

[12] U.S. Department of Defense. Department of Defense Strategy for Operating in Cyberspace. https://csrc.nist.gov/CSRC/media/Projects/ISPAB/documents/DOD-Strategy-for-Operating-in-Cyberspace.pdf, July 2011. (Accessed on 01/22/2022).

[13] 川口貴久. サイバー空間における安全保障の現状と課題 – サイバー空間の抑止力と日米同盟 –. https://www2.jiia.or.jp/pdf/resarch/H25_Global_Commons/03-kawaguchi.pdf, March 2014. (Accessed on 01/22/2022).

[14] 防衛省・自衛隊. 令和 3 年版防衛白書 | 3　サイバー空間における脅威に対する取組. https://www.mod.go.jp/j/publication/wp/wp2021/html/n130303000.html, September 2021. (Accessed on 01/22/2022).

[15] Anti-Phishing Working Group (APWG). Phishing Attack Trends Report – 2nd Quarter 2021, June 2021.

[16] Elmer E.H. Lastdrager. Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science*, Vol. 3, No. 9, pp. 1–16, 2014.

[17] Kaspersky. What is Social Engineering? https://www.kaspersky.com/resource-center/definitions/what-is-social-engineering, 2017. (Accessed on 01/22/2022).

[18] Koceilah Rekouche. Early Phishing. pp. 1–9, June 2011.

[19] Kang Leng Chiew, Kelvin Sheng Chek Yong, and Choon Lin Tan. A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications*, Vol. 106, pp. 1–20, September 2018.

[20] フィッシング対策協議会技術・制度検討ワーキンググループ. フィッシングレポート 2021. https://www.antiphishing.jp/report/phishing_report_2021.pdf, June 2021. (Accessed on 01/23/2022).

[21] NTT ドコモ. ドコモからのお知らせ：「危険 SMS 拒否設定」（無料）の提供について. https://www.nttdocomo.co.jp/info/notice/page/220113_00.html?icid=CRP_INFO_anti-phishing_to_CRP_INFO_notice_page_220113_00, January 2022. (Accessed on 01/23/2022).

[22] ソフトバンク. 迷惑 SMS 対策機能を提供開始｜プレスリリース｜ニュース｜企業・IR｜. https://www.softbank.jp/corp/news/press/sbkk/2022/20220113_02/, January 2022. (Accessed on 01/23/2022).

[23] Ron Amadeo and Ars Technica. Take one last look at Google Toolbar, which is now dead. https://arstechnica.com/gadgets/2021/12/happy-21st-birthday-to-google-toolbar-which-inexplicably-still-exists/, December 2021. (Accessed on 01/09/2022).

[24] 窓の杜. ヤフー、「Yahoo!ツールバー」のサービス終了を発表. https://forest.watch.impress.co.jp/docs/news/1071441.html, July 2017. (Accessed on 01/09/2022).

[25] Min Wu, Robert C. Miller, and Simson L. Garfinkel. Do Security Toolbars Actually Prevent Phishing Attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pp. 601–610, New York, NY, USA, April 2006. Association for Computing Machinery.

[26] Aiping Xiong, Robert W. Proctor, Weining Yang, and Ninghui Li. Is Domain Highlighting Actually Helpful in Identifying Phishing Web Pages? *Hum Factors*, Vol. 59, No. 4, pp. 640–660, June 2017.

[27] Christopher Thompson, Martin Shelton, Emily Stark, Maximilian Walker, Emily Schechter, and Adrienne Porter Felt. The Web's Identity Crisis: Understanding the Effectiveness of Website Identity Indicators. In *Proceedings of the 28th USENIX Conference on Security Symposium*, SEC'19, pp. 1715–1732, USA, August 2019. USENIX Association.

[28] フィッシング対策協議会. 利用者向けフィッシング詐欺対策ガイドライン 2020 年度版. https://www.antiphishing.jp/report/consumer_antiphishing_guideline_2020.pdf, June 2020. (Accessed on 01/15/2022).

[29] Sara Albakry, Kami Vaniea, and Maria K. Wolters. *What is This URL's Destination? Empirical Evaluation of Users' URL Reading*, pp. 1–12. Association for Computing Machinery, New York, NY, USA, April 2020.

[30] 月浦崇, 定藤則弘. 想起・誤想起（記憶）– 脳科学辞典. https://bsd.neuroinf.jp/wiki/%E6%83%B3%E8%B5%B7%E3%83%BB%E8%AA%A4%E6%83%B3%E8%B5%B7%EF%BC%88%E8%A8%98%E6%86%B6%EF%BC%89, February 2013. (Accessed

on 01/20/2022).

[31] W. A. Bousfield, J. Esterson, and G. A. Whitmarsh. The Effects of Concomitant Colored and Uncolored Pictorial Representations on the Learning of Stimulus Words. *Journal of Applied Psychology*, Vol. 41, No. 3, pp. 165–168, June 1957.

[32] Allan Paivio and Kalman Csapo. Picture Superiority in Free Recall: Imagery or Dual Coding? *Cognitive Psychology*, Vol. 5, No. 2, pp. 176–206, September 1973.

[33] Douglas L. Nelson, Valerie S. Reed, and Cathy L. McEvoy. Learning to Order Pictures and Words: A Model of Sensory and Semantic Encoding. *Journal of Experimental Psychology: Human Learning and Memory*, Vol. 3, No. 5, pp. 485–497, September 1977.

[34] Brandon A. Ally. Using Pictures and Words To Understand Recognition Memory Deterioration in Amnestic Mild Cognitive Impairment and Alzheimer's Disease: A Review. *Current Neurology and Neuroscience Reports*, Vol. 12, No. 6, pp. 687–694, December 2012.

[35] Rebecca G. Deason, Neil A. Nadkarni, Michelle J. Tat, Sean Flannery, Bruno Frustace, Brandon A. Ally, and Andrew E. Budson. The Use of Metacognitive Strategies to Decrease False Memories in Source Monitoring in Patients with Mild Cognitive Impairment. *Cortex*, Vol. 91, pp. 287–296, June 2017.

[36] Marilyn A. Borges, Mary Ann Stepnowsky, and Leland H. Holt. Recall and Recognition of Words and Pictures by Adults and Children. *Bulletin of the Psychonomic Society*, Vol. 9, No. 2, pp. 113–114, February 1977.

[37] William E. Hockley. The Picture Superiority Effect in Associative Recognition. *Memory & Cognition*, Vol. 36, No. 7, pp. 1351–1359, October 2008.

[38] Allan Paivio. *Imagery and Verbal Processes*. Psychology Press, November 1979.

[39] Fergus I.M. Craik and Robert S. Lockhart. Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior*, Vol. 11, No. 6, pp. 671–684, December 1972.

[40] Lionel Standing. Learning 10,000 Pictures. *The Quarterly Journal of Experimental Psychology*, Vol. 25, No. 2, pp. 207–222, May 1973.

[41] Robert E. Gehring, Michael P. Toglia, and Gregory A. Kimble. Recognition Memory for Words and Pictures at Short and Long Retention Intervals. *Memory & Cognition*, Vol. 4, No. 3, pp. 256–260, May 1976.

[42] Phillip Isola, Devi Parikh, Antonio Torralba, and Aude Oliva. Understanding the Intrinsic Memorability of Images. In J. Shawe-Taylor, R. Zemel, P. Bartlett,

F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 24. Curran Associates, Inc., 2011.

[43] R. Jenkins, A. J. Dowsett, and A. M. Burton. How Many Faces Do People Know? *Proceedings of the Royal Society B: Biological Sciences*, Vol. 285, No. 1888, p. 1319, October 2018.

[44] Janine Willis and Alexander Todorov. First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, Vol. 17, No. 7, pp. 592–598, July 2006. PMID: 16866745.

[45] Peter J.B. Hancock, Vicki Bruce, and A.Mike Burton. Recognition of Unfamiliar Faces. *Trends in Cognitive Sciences*, Vol. 4, No. 9, pp. 330–337, September 2000.

[46] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *Communications of the ACM*, Vol. 63, No. 11, pp. 139–144, November 2020.

[47] Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. Lexical Feature Based Phishing URL Detection Using Online Learning. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security*, AISec '10, pp. 54–60, New York, NY, USA, October 2010. Association for Computing Machinery.

[48] Rami M. Mohammad, Fadi Thabtah, and Lee McCluskey. Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications*, Vol. 25, No. 2, pp. 443–458, 2014.

[49] Mohammed Nazim Feroz and Susan Mengel. Phishing URL Detection Using URL Ranking. In *2015 IEEE International Congress on Big Data*, pp. 635–638, New York, NY, USA, August 2015.

[50] Ankesh Anand, Kshitij Gorde, Joel Ruben Antony Moniz, Noseong Park, Tanmoy Chakraborty, and Bei-Tseng Chu. Phishing URL Detection with Oversampling based on Text Generative Adversarial Networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 1168–1177, December 2018.

[51] Ahmed AlEroud and George Karabatis. Bypassing Detection of URL-Based Phishing Attacks Using Generative Adversarial Deep Neural Networks. In *Proceedings of the Sixth International Workshop on Security and Privacy Analytics*, IWSPA '20, pp. 53–60, New York, NY, USA, March 2020. Association for Computing Machinery.

[52] Alexa Internet. Alexa - Top sites, 2022. (Accessed on 01/20/2022).

[53] U.S. Department of Education and National Center for Education Statistics. Fast Facts: Students with disabilities. `https://nces.ed.gov/fastfacts/display.asp?`

id=64, 2021. (Accessed on 01/15/2022).

[54] 発達性ディスレクシア研究会. ディスレクシアを理解するために, April 2014.

[55] Akira Uno, Taeko N. Wydell, Noriko Haruhara, Masato Kaneko, and Naoko Shinya. Relationship between reading/writing skills and cognitive abilities among Japanese primary-school children: Normal readers versus poor readers (dyslexics). *Reading and Writing*, Vol. 22, No. 7, pp. 755–789, August 2009.

[56] 中嶋智史, 請園正敏, 須藤竜之介, 布井雅人, 北神慎司, 大久保街亜, 鳥山理恵, 森本裕子, 高野裕治. 日本語版 20 項目相貌失認尺度の開発および信頼性・妥当性の検討. 心理学研究, Vol. 90, No. 6, pp. 603–613, February 2020.

[57] Bitmain. ANTMINER S19 XP. https://shop.bitmain.com/product/detail?pid=00020211207113044632h8EJMWk30658. (Accessed on 01/16/2022).

発表文献

[1] 山崎慎治, 宮本大輔. 顔画像生成技術を用いた偽造ウェブサイト判別支援手法の提案. 2021 年暗号と情報セキュリティシンポジウム予稿集（SCIS2021）, 3C3-4, January 2021.（オンライン）.

[2] 山崎慎治, 宮本大輔. 偽造ウェブサイトに対する顔画像生成技術を用いた判別支援手法の設計と実装. コンピュータセキュリティシンポジウム 2021 論文集（CSS2021）, 1D4-2, October 2021.（オンライン）.

謝辞

「サイバーセキュリティの研究がしたい」との希望が叶い，研究室に配属されてから 2 年が経過しました．コロナ禍という未曾有の事態において，2 年の間に本学構内に足を踏み入れたのは片手で数えるほど，また研究室には終ぞ入室することなく，かつて思い描いていた修士学生生活とはかけ離れたものでした．思ったように研究に取り組むことができなかった状況でしたが，そんな私を見捨てることなく，周囲の方々にご助力いただくことで本論文の執筆に至りました．ここに心より感謝を申し上げます.

指導教員である宮本大輔准教授には 2 年間の研究室生活において，研究内容から日々の学生生活の心構えに至るまで，厳しくも暖かいご指導ご鞭撻をいただきました．通学が難しい私の事情を汲んで事務手続き等で融通していただき，コロナ禍の異常事態でも私が人とのつながりを築けるように，先生の紹介で様々な人とオンラインで面会が叶いました．また私が研究者として極めて未熟であったにもかかわらず，一つ一つ時間を割いて丁寧に指導いただき，時には議論を，時には的確なアドバイスをいただきました．先生なくして私の修士研究はありませんでした．心より感謝いたします.

奈良先端科学技術大学院大学 門林雄基教授には，ご多忙の中で研究についてのアドバイスを頂きました．深く広い経験に基づいた指摘は勘所を押さえており，本研究を深めるにあたって非常に参考になりました．深く感謝申し上げます.

同分野で研究を行っている同期の祐村昌秀氏には多くの助言をいただきました．特に開室して 1 年目の研究室で，私以外に学生のいなかった研究室では，祐村氏の存在は欠かせないものでした．深く感謝します．伊藤彰秀氏，井口和真氏，高橋大成氏を始め，折りに触れ日々の息抜きに付き合っていただいた同期の皆様に感謝します.

制限された日々の生活の中でも，刺激を与え精神面で支えてくださいました大空スバル氏，星街すいせい氏，大神ミオ氏に深く感謝します.

最後に，あらゆる面からこれまでの学生生活を支えてくださった家族，友人，先生，そして全ての方々に深く感謝いたします．本当にありがとうございました.

付録 A

調査に使用した画像

本章では調査時に実験参加者に提示し，記憶または判別してもらった画像を示す.

a: 提案群向けに提示

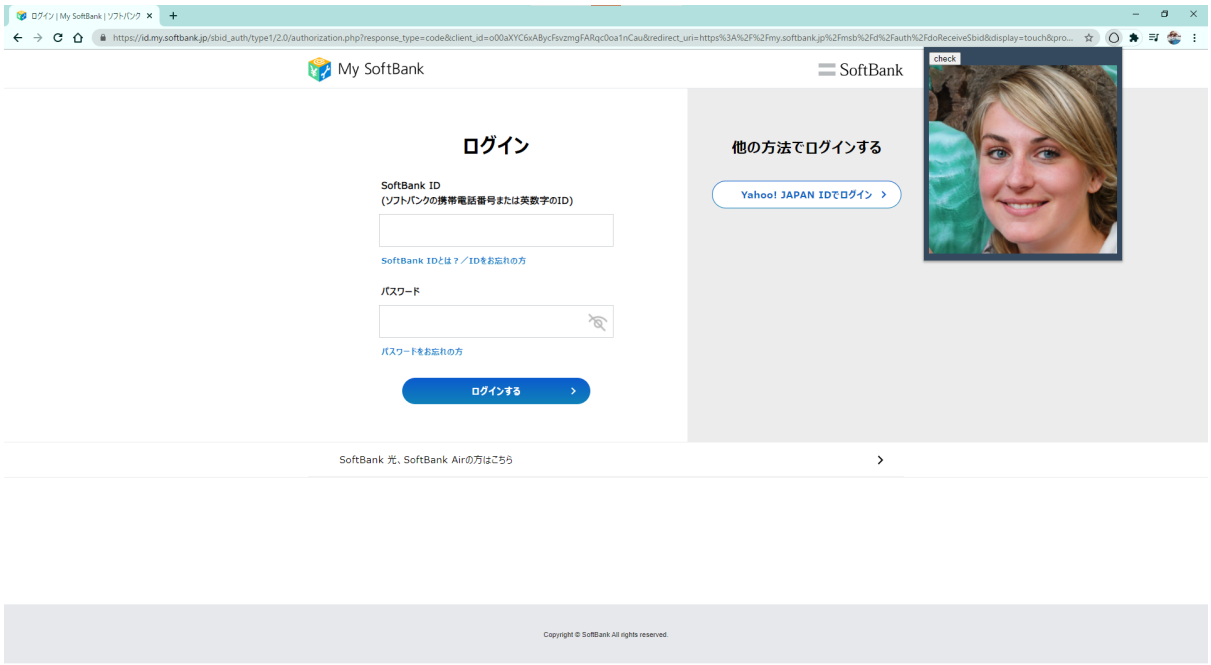
b: 対照群向けに提示

図 19　1 問目に提示したフィッシングサイトの画像

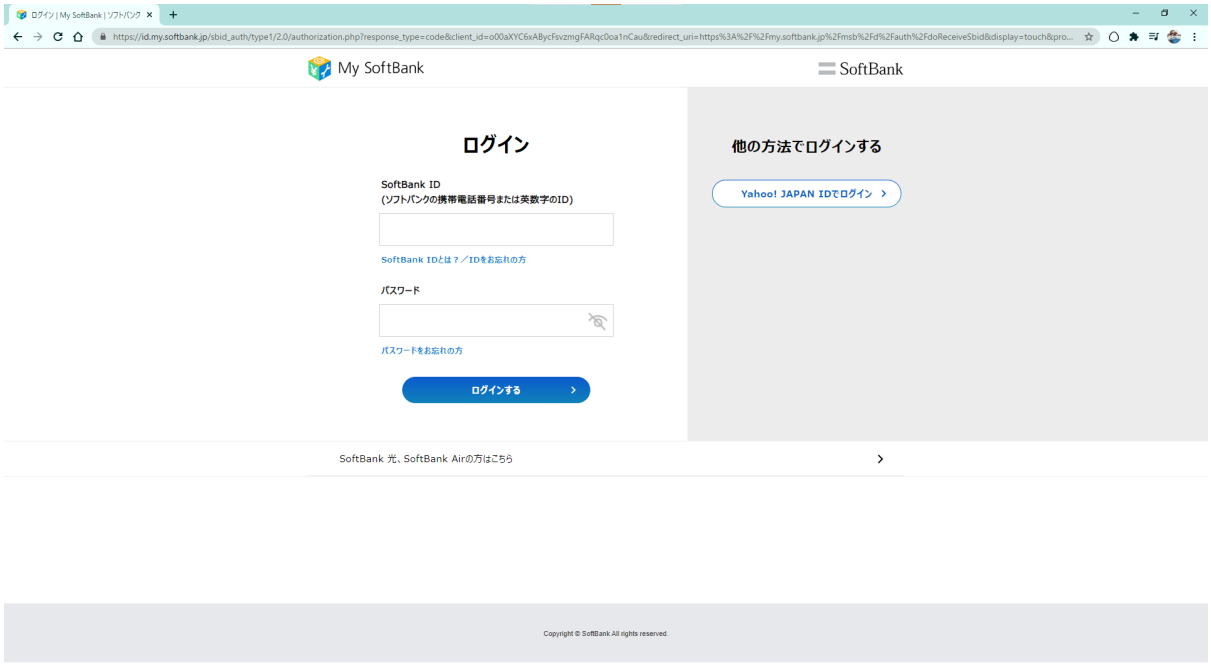
a: 提案群向けに提示

b: 対照群向けに提示

図 20　2 問目に提示したフィッシングサイトの画像



a: 提案群向けに提示



b: 対照群向けに提示

図 21　3 問目に提示した正規のウェブサイトの画像

a: 提案群向けに提示

b: 対照群向けに提示

図 22　4 問目に提示した正規のウェブサイトの画像

a: 提案群向けに提示

b: 対照群向けに提示

図 23　5 問目に提示したフィッシングサイトの画像

a: 提案群向けに提示

b: 対照群向けに提示

図 24　6 問目に提示したフィッシングサイトの画像

a: 提案群向けに提示

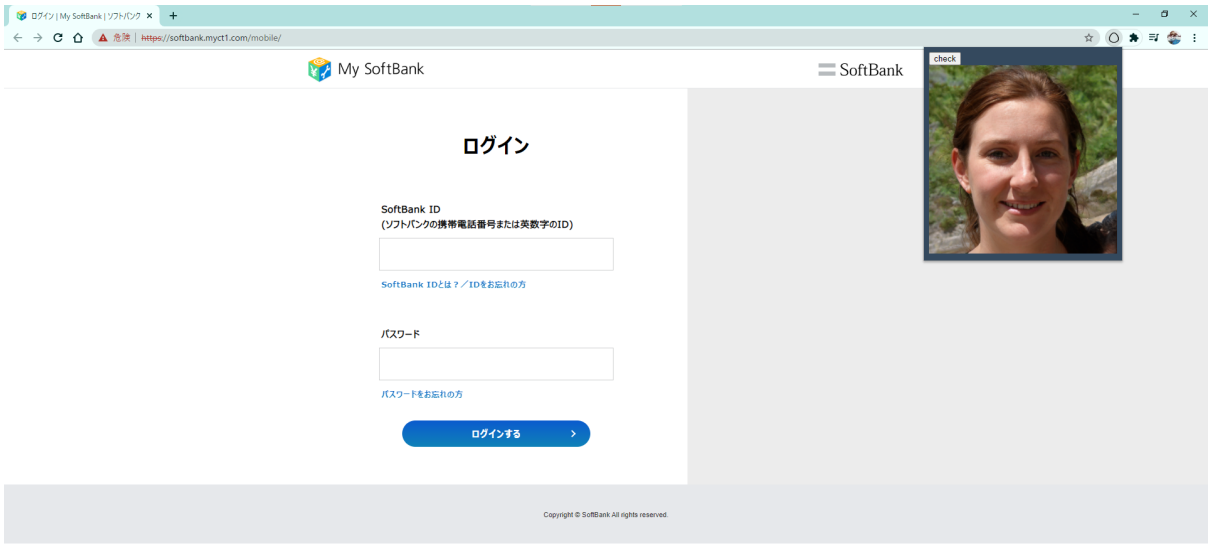
b: 対照群向けに提示

図 25　7 問目に提示した正規のウェブサイトの画像

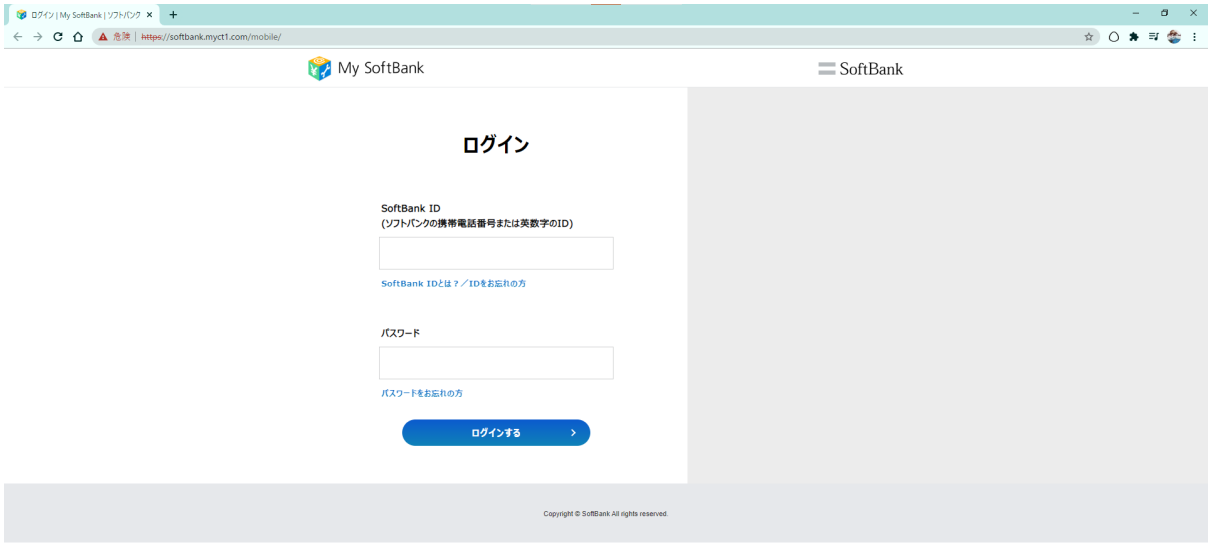
a: 提案群向けに提示

b: 対照群向けに提示

図 26　8 問目に提示したフィッシングサイトの画像



a: 提案群向けに提示



b: 対照群向けに提示

図 27　9 問目に提示したフィッシングサイトの画像

a: 提案群向けに提示

b: 対照群向けに提示

図 28　10 問目に提示した正規のウェブサイトの画像