

Web ページ上のクリック操作にまつわる BADUI の VLM を用いた改善手法

徳原 真彩^{1,a)} 木下 裕一朗¹ 中村 聡史¹

概要: 企業や各種団体が Web サイトやアプリケーションなどを用いてサービス提供を行う事が一般的になってきたが、分かりづらいインタフェースも存在しており日々、人を困らせている。そこで本研究では、VLM が人のように使いにくいインタフェースを誤認することに着目し、Web サイトにおいてクリックが可能な要素の中で、ユーザがクリックできないと認知する可能性があるものを VLM を用いて推定し、Web ページのユーザへと強調表示により提示する手法を提案した。また Chrome の拡張機能として実装し、システムを用いて効果を検証した。システムを用いた評価の結果、クリック可能要素全体のうち約 7 割の要素を認識できることが分かった。また、UI の使いにくさに対して VLM がどのように判定するかについての検討も行った。

1. はじめに

多くの企業や団体が、情報発信やサービス提供を目的として、Web サイトやアプリケーションなどを活用しており、それらの利用は一般的となってきた [1]。また、PC・スマートフォンの普及により Web サービスの利用者は増加し、利用者層の多様化や Web サービスの多機能化などが進んでおり、多くの企業は自社の商品やサービスを魅力的に見せるため、工夫を凝らしたデザインを取り入れる事が多い。さらに様々なフレームワーク等が登場したことで、サイトのデザインはより複雑で凝ったものになっている。しかし、UI に問題を抱えたサイトも多く、実際に利用する際に目的のページにたどり着くのに時間がかかったり、分かりやすい誘導ができていないためにユーザの意図していないページへ遷移させてしまったりすることがある。また、独自性の高いデザインはユーザにとって見慣れない UI となり、思った通りの動作を実行できない可能性もある。

サービスを行う Web サイトにおいては、初めてそのサイトを訪問するユーザであっても瞬時にそれらの認知と操作を考えると無く利用できることが望ましい。広川ら [2] は、直感的インタフェースの特徴としてユーザが操作対象を意識的な意味解釈をしなくても瞬時に認知し操作できることを挙げている。こうした UI を実現するため、UI 開発者が参照するために作られた UI ガイドライン [3][4] があ



図 1: 上 5 つの画像はクリックできず、ページ下部の「ネットゼロ 5 分野の取り組み」がクリック可能なサイト



図 2: 「購入申込」はクリックできず、「購入申込みはこちら」がクリック可能であるアプリケーション

るが、ガイドラインを参照することで分かりやすい UI の特徴を知ることができるものの、実際に開発した UI の大きさや配置、色合いに対してユーザがどのように認知するかを、UI の開発段階で予想するのは難しい。また、システムの開発者はその UI に慣れすぎているため、その問題に気づきづらいという問題がある。

図 1 と図 2 にユーザの利用の際に時間を要してしまう可

¹ 明治大学
Meiji University
^{a)} tokuhara@nkmr-lab.org

能性のある UI の例を示す。図 1 に示すサイトは、上の 5 分野を示すイラスト群はクリック不可能であり、下の「ネットゼロ 5 分野の取組み」を押さなければ詳細が確認できない。この 5 分野のイラスト群（クリック不可）が、クリック可能性を示すシグニファイア [5] として機能してしまうため、無駄なクリック操作が何度も行われてしまうなど、問題のある UI といえる。また、図 2 に示す例は、ユーザが購入申し込みを目的として利用するものであるが、青色で装飾されている「購入申込」は操作できない。ここでは「購入申込みはこちら」の部分进行操作する必要があるが、操作可能性を示す矢印が文字列と重なっているため、操作可能であると判断される可能性が低くなっているものである。こうした作成した UI に対してデザイナーにフィードバックを行う関連研究は多く存在する [6][7]。しかし、既にリリースされているサービスに対してユーザが用いる UI 改善ツールについては十分に研究されていない。

そこで本研究では、こうしたデザインによる不便さを解消するため、ユーザがブラウザでサイトを利用する際に、ユーザが認知しづらいクリック可能位置を検出し、その要素を分かりやすくユーザに示すことによってサイトの使いづらさを軽減する Chrome 拡張機能を作成し、検出精度を測ることを目的とする。また、作成したツールを用いて、使いづらいと感じるユーザが多いとみられるサイトを対象に有用性があると感じられる例を挙げる。

2. 関連研究

2.1 UI が与える影響

UI が人々の認知や行動、評価に与える影響についての研究は様々ある。

Tichindelean ら [8] は、Web サイトのユーザビリティを分析した結果、情報の配置や色彩などのデザイン要素がユーザの視線行動と情報の記憶に大きく影響を与えたことを示した。また Riegler ら [9] は、モバイルアプリケーションの UI の複雑性を定量化し、UI 要素の数や密度、色の組み合わせなどの視覚的要素がユーザの使いやすさや作業負担にどの程度影響を与えるか調査し、UI の複雑さが高いほど作業効率が低下し、ユーザビリティが低く評価されることを明らかにした。Sundar ら [10] はクリックやスライド、ズームなどを含む 6 つの操作手法がユーザ体験に与える影響を分析し、スライドは記憶力を向上させ、ズームはユーザの評価が低い傾向を確認している。また、直感的で自然な操作手法がユーザ体験を向上させる一方で、これらの手法を組み合わせた複雑な操作は評価を下げる可能性を示した。また、久保 [11] は「UX のハニカム構造」に含まれる、使いやすさや探しやすさが、EC サイトの利用においてブランド態度へ正の影響を与えることを明らかにしている。

これらのことから、操作が直感的であると感じる要素には UI のデザイン性も含まれていると考えられるため、ユー

ザ体験の向上には UI が重要なことが分かる。また、UI がユーザ体験に影響を与え、それらを扱うブランドへの印象も変化させる可能性があるため、ユーザが利用しやすい UI を提供することが重要である。

2.2 UI の視覚的評価

作成した UI に対する視覚的な評価を得るシステムはリリース前にデザインの改善を行う際の手掛かりとなり、様々な手法が提案されている。

Duan ら [6] は VLM (Vision-Language Model) を用いて、UI デザインとガイドラインを入力することによって UI のフィードバックを行うシステムを構築し、微細なエラーの発見やテキスト改善に有用性があることを示している。また、Bisante ら [12] は Web インタフェースのユーザビリティ評価を支援するツール「CWGPT」を開発した。また実験により、本ツールがユーザが利用しづらい UI を検出し、デザイナー初心者にとって有用であり、インタフェース設計スキルの向上に寄与する可能性があるとしている。Deka ら [13] はモバイルアプリのデザイン評価を効率化する「ZIPT」を提案しており、既存の Android アプリケーションに対して、ソースコードにアクセスすること無くユーザ操作データを収集し、視覚化することで、デザインのパフォーマンスや改善点を特定することを可能にしている。Eldon ら [14] は、Web ページ上でユーザが特定の要素を見つけるまでの時間を高精度で予測するモデルを作成した。またモデルは画像データとターゲットの特徴を組み合わせた予測を行い、予測の中で生成される注目されやすい要素である「Saliency Heatmap」を提示し、デザインの改善に役立てることができることを明らかにしている。

一方、現在ある UI に対して、類似の UI の生成や検索を行うモデルも提案されている。Wu ら [15] はスクリーンショットから UI モデルを解析する「Screen Parsing」というシステムを開発した。これは要素の検出や、その階層構造の予測、要素グループの分類を行うシステムであり、類似 UI の検索や UI コード生成といった応用に利用可能であるとしている。また Wu ら [16] は、スクリーンショットと自然言語の説明を基に UI の品質を評価するモデル「UIClip」を提案し、数値スコアでの評価やデザイン改善の提案を提供するとともに、UI コードの生成や UI 例の検索などの応用も行っている。

このように UI に対して視覚ベースの評価を行い、結果をデザイナーにフィードバックすることによってデザインの改善を促す研究が行われているが、本研究では、利用しづらい UI をユーザに対してフィードバックすることを目的としている。

2.3 UI テストの自動化

サイトやアプリケーションにおける目的の遂行が、現在

の UI から容易に行えるかを評価する研究も行われている。

Gao ら [17] は、GUI 操作の自動化を評価するためのベンチマーク「ASSISTGUL」を提案し、100 のタスクを用いてモデルの性能を検証した結果、提案手法が従来よりも高い成功率になることを示している。Ran ら [18] は、VLM を活用した自動 UI テストを行う「Guardian」を提案し、利用のプロセスにおいて動的な再計画を行う事によって成功率を高めることを明らかにした。また Feiz ら [19] は、スマートフォンアプリケーションのスクリーンショットを用い、画面の類似性と遷移を理解するための 2 つの機械学習モデルを提案することで、クロールやテストの自動化に応用可能であることを明らかにした。さらに Dongping ら [20] は、動的なタスクを含む多様な UI をカバーしたデータセットを用いた新モデル「GUI-Vid」を提案することにより、UI 操作の理解能力向上を達成している。

これらの研究は、目的が達成されるまでの一連の流れの中における問題点を抽出し、指摘を行うものであるが、本研究ではある一場面におけるクリック可能性に着目し、その問題点をユーザへと通知するシステムを提案する。

2.4 ユーザの行動予測

UI に対してユーザがどのような印象を持つのかを推測する研究も存在している。

Swearngin ら [7] はモバイルインタフェースにおいて、ユーザが要素をタップ可能と認識する確率を予測する深層学習モデルを開発し、その精度は 90.2% と高い評価を得ていることを示した。また、Wu ら [21] は「Never-ending UI Learner」を提案している。これは、システムがアプリケーションをクロールし、「タップ可能性」等を学習するモデルであり、既存の手手でアノテーションを行ったデータセットを越える精度を示し、長期間にわたるデザインの進化にも対応することができるとしている。Yuan ら [22] は Web ページ上での視覚的検索性能を予測するための深層学習モデルを開発した。このモデルは、画像データやターゲットの特徴を組み合わせ、ユーザが特定の要素を見つけるまでの時間を高精度で予測する。Zhou ら [23] はモバイルデバイス上のクリック行動を深層学習を用いてモデル化し、ユーザのクリック履歴や UI の構造情報などから次にクリックされる可能性の高い UI 要素を予測する手法を提案した。Bylinskii ら [24] は人間のクリックデータや重要性を評価から、デザイン要素の中でどの部分が視覚的に重要かを自動で判別できるモデルを開発した。

これらの研究は、実際のユーザの行動予測をデザイナーにフィードバックすることで客観的なデータを利用したデザインの改善を支援するものである。本研究では、ユーザの行動予測を拡張機能に適応させることで、クリック可能予測が難しい要素をユーザに提示するような支援を目指す。

3. 提案手法と実装

3.1 本研究で対象とする UI と VLM による判定

ユーザがクリック可能な位置をクリック可能であると認知できなければ、ユーザはその要素を見逃してしまい、目的のページにたどり着くことや、操作に時間を要してしまう。ここで、サイト上のクリックに関する要素は表 1 のように分類できると考えた。

表 1 において、(1) と (4) はユーザの認知とシステムの挙動が一致しているため問題が無いが、(2) と (3) はユーザの認知とシステムの挙動のずれが存在するため、ユーザの利用を妨げる UI であると言える。具体例を挙げると、(2) は図 1 のようなものであり、(3) は図 2 のようなものである。ここで (3) は特に、一度クリックできないと思い込んでしまったうえ、他の位置に同じリンクを発見できなかった場合は、サイト全体から再度探し直さなければならないためより一層不便を強いる UI となってしまう。そこで本研究での提案手法では、Chrome の拡張機能を用いて、まずはサイト上の (3) をユーザに示す機能の実現を目指す。

ユーザがクリックできると認知するかどうかについては機械学習など様々な方法が考えられるが、VLM が BADUI に遭遇した時に、ひとと同じように間違ってしまうこと [25] に注目し、ここでは画像を認識可能な VLM を利用する。実際に、図 2 について、VLM に購入する際に優先的にクリックすべき箇所はどこかというプロンプトで質問したところ、まず目立つ「購入申込」が提示され、次に「購入申込みはこちら」が挙げられていた。

3.2 提案する Chrome 拡張機能

本研究で提案する Chrome 拡張機能の動作イメージを図 3 に示す。以下に示す順に処理を行った上で、ページ閲覧中のユーザがクリック可能だと認知することが難しい要素を示す。なお、本研究において実際にはクリック可能であるにも関わらず、VLM がクリック可能だと予測しなかった要素は、ユーザにとってもクリック可能だと認知することが難しい要素であると仮定した。また、VLM を用いた理由は、クリックできそうであるという推測は人と VLM で似たような考えで行われていると考えたためである。

- 閲覧中のページのスクリーンショットを撮影し、VLM へとクリック可能位置を予測し、返答するプロンプトと共に画像を送信
- 拡張機能において、送信したスクリーンショットの中から実際にクリック可能な要素を抽出

表 1: クリック要素分類

	クリック可能要素	クリック不可能要素
クリックできそうな要素	○ (1)	△ (2)
クリックできなさそうな要素	× (3)	○ (4)

表 2: 文字部分特定と画像部分特定のプロンプト

文字部分特定のプロンプト	画像部分特定のプロンプト
<p>あなたは画像内のリンクやクリック可能な要素を特定するために使用されます。以下の手順に従ってください。</p> <ol style="list-style-type: none"> 1. クリック可能なリンクまたは要素と思われる箇所を特定してください。 <ul style="list-style-type: none"> - 下線付きの文字列 - ボタンのようなデザインの要素 - 画像でリンクとして使用されているもの（例えばアイコンやバナー画像） - 見た目がクリック可能に見える要素 2. 各クリック可能な箇所の文字列を以下の形式で正確に出力してください： ["string": "詳しくはこちら"] 3. 出力する際には、JSON 形式を正確に守ってください。コメントや説明文を含めないでください。 4. **重要** クリック可能性が少しでもありそうな要素は全てリストしてください。この形式を守り、正確な出力を生成してください。 	<p>あなたは画像内のリンクやクリック可能な要素を特定するために使用されます。以下の手順に従ってください。</p> <ol style="list-style-type: none"> 1. 画像でリンクとして使用されているもの（例えばアイコンやバナー画像）と思われる箇所を特定してください。 <ul style="list-style-type: none"> - お手続きはこちらなどのようにクリックができることが示唆される文が入っている - 企業のロゴや SNS のロゴなどクリックできる対象となることが多い画像 2. 各クリック可能な箇所を以下の形式でどこを指しているのかが分かるように正確に出力してください： ["string": LINE のロゴ, "string": 企業のロゴ, "string": お問い合わせと書かれた画像, "string": 赤いバッグの商品の画像, "string": エビが 5 尾写った画像] 3. 出力する際には、JSON 形式を正確に守ってください。コメントアウトによるコメントや説明文を含めないでください。 4. **重要** クリック可能性が少しでもありそうな要素は全てリストしてください。この形式を守り、正確な出力を生成してください。

- 実際にはクリック可能な要素であるが、VLM がクリック可能だと予測しなかった要素を抽出
- 抽出された差分の要素のみ強調表示を行い、クリック可能位置を分かりやすく提示

なお、クリック可能位置を抽出し、その全てを強調表示しユーザへと示した場合、サイト上の全てのクリック可能位置が強調表示されることでサイトの見た目が大きく変わってしまい、デザイナーが意図したデザインと離れてしまう可能性がある。そのため、認知の難しい要素だけを示す拡張機能を目指した。

3.3 拡張機能の実装

クリック可能位置の予測には OpenAI 社が提供する、VLM の ChatGPT API を利用しモデルは gpt-4o とした。また、使用したプロンプトを表 2 に示す。一般的なシグニファイアなどの特徴は説明しつつ、例を交えたプロンプトとした。

図 4 にシステムの利用イメージを示す。拡張機能で画面

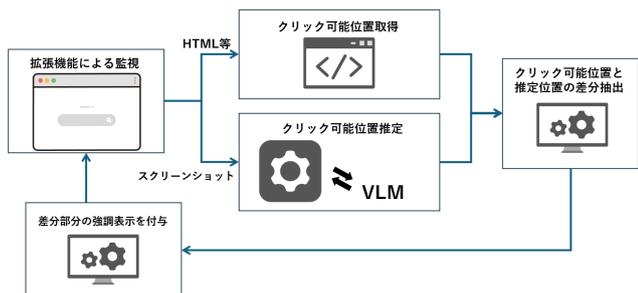


図 3: Chrome 拡張機能イメージ図

内のクリック可能要素を抽出したうえで、その上に半透明の四角形を表示することでクリックが可能であることを示している。

4. 手法の精度評価

4.1 評価方法

本稿では、サイト内のクリック可能位置のうち、VLM がクリック可能だと判定した要素がどの程度除外できるかを測るために、作成した Chrome 拡張機能を用いた検出率の評価を行った。評価対象としては拡張機能を適用することが可能であったサイトの内、日本でのアクセス数が上位 30 のサイト [26] を選定した。なお、ページ全体のスクリーンショットを用いると、要素の量によって VLM の返答量の上限に達してしまう可能性があるため、今回の評価ではスクリーン (1504 × 842) に表示できる領域のスクリーンショット 1 枚を用いた。また、サイトのスクリーンショットの位置は、各項目へのリンクが密集していることが比較

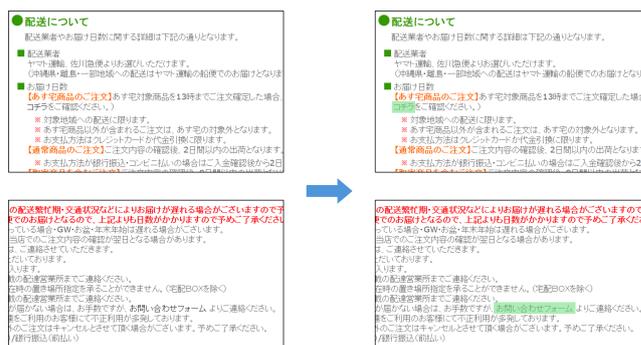


図 4: システム利用イメージ図

的多いページ上部とした。

サイト内のクリック可能位置は大きく文字部分と画像部分の2つに分けられる。今回の評価では文字部分と画像部分に分けて、各3回ずつVLMにクリック可能位置の予測を行ってもらい、その結果を繋ぎ合わせたものをVLMがクリック可能だと予測したものとした。

また本評価では、クリック可能な要素のうち、他のページへと遷移することのできるリンクを対象とした。クリックするとメニューやポップアップが出現する要素、ホバーすることによって新たなメニューが出現する要素、メニューを切り替えるタブの要素等は対象外とした。なお、今回は文字列部分の強調表示の除外はシステムにおいて可能であったが、画像部分の対応付けが難しかったため、評価は手動で行った。

4.2 検出率

本稿では、クリック可能であるがクリック可能と推測されなかった場合を未認識、クリック不可能であるがクリック可能と推測された場合を誤認識と呼ぶ。未認識率を「実際にはクリック可能であるがVLMがクリック可能と予測しなかった要素の数」を「実際にクリックが可能な要素の数」で割った割合として求めた結果、約29%（全要素数1075のうち314）となった。ここで求めた未認識率は表1の(3)の割合にあたる。未認識率は約3割となり、全体の要素のうち約7割の要素を省き、明らかなクリック可能箇所を除いた強調表示をすることが可能であることが分かった。

次に図5に各サイトごとの未認識率を示す。全体での未認識率は約3割であったが、半数以上は2割を下回っている。未認識率が5割を超えるサイトが6つあったため、中央値は15%であった。このことより、顕著に高い未認識率となってしまったサイトの特徴を捉えたプロンプトに修正することによって未認識率は下げることが可能であると考えられる。

また、誤検出の例としては表3のような特徴が見られた。

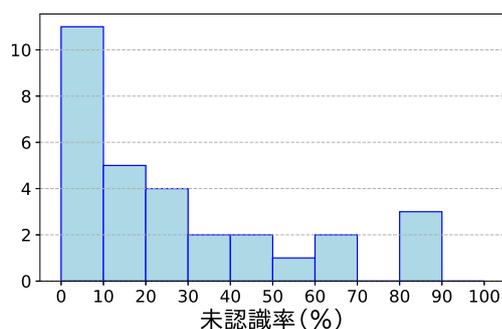


図5: サイトごとの未認識率の分布

5. 考察

5.1 誤検出と未認識率の捉え方

本研究の提案手法においては、誤検出は機能上問題は起こらない。しかし、誤検出が多いサイトが存在した場合、ユーザにとってクリック可能だと見間違えてしまう要素が多く、利用しづらいサイトだと言える。今回評価対象としたサイトはアクセス数が多く、長期間の運営によりフィードバックを得ながら改良されてきたデザインであると考えられるため、誤検出はあまり見られなかったが、リリース直後のサイト等は今回の評価より誤検出や未認識が発生する可能性は高くなる。クリックが可能だという誤認が起これば、より多くの要素に対してクリックする動作を起こさなければならないため、ユーザ体験は低下すると考えられる。そのため、誤検出が多いサイト等においてはクリック可能位置をより目立たせるなどの対策を取ることでもユーザ体験の向上に寄与する可能性がある。そこで今後は、実験協力者に実際にシステムを構築してもらい、その構築したシステムに対してシステムが機能するかを検証する予定である。

未認識率は全体の要素のうち強調表示を行う箇所の割合に該当する。本研究の目的は人々がクリック可能だと認知することが難しい要素の検出であるため、未認識率が極端に高くなると強調表示が増えてしまい、分かりづらい要素を示すシステムとは言えなくなってしまう。そのため、今後は極端に未認識率が高くなってしまふサイトを対策したプロンプトやシステムの構築をする必要があると考えられる。

5.2 未認識が起きた要素

本節では、実際に未認識が発生した要素の実例を挙げる。

まず、ブログの投稿サイトや動画投稿サイトなどのタイトルやユーザネームについて述べる。これらはそれぞれのユーザが決定するものであり、クリック可能性を示した文言ではない。また、その記事や動画のサムネイルについても投稿者自身が決定する要素のため、一部の画像はクリック可能性を示していない。

図6に示すニュース記事一覧において、文字の色を青色に変更したクリック可能性の提示が行われているが、記事のタイトル部分の要素が見逃されてしまっていた。こちらでも同様に文字の内容自体がクリック可能性を示さないことから、クリック可能性を示していると判断されず抽出されなかった可能性がある。しかし、これらの要因を含む要素は一般的にクリック可能であることがユーザの間で認知されていると考えられる。そのため、ユーザがどの程度認知可能か調査を行う必要はあるが、プロンプトやシステムによる除外を行うことが適切であると考えられる。

表 3: 誤検出や未認識となった要素の特徴例

誤検出の例	未認識の例
<ul style="list-style-type: none"> 他の文字と比較すると太文字で書いてあり、「関東」などのように一覧に飛べる可能性の高い単語 「～で探す」などの検索に関わる動詞で終わっていてもクリックできない文字 他にクリック可能要素があるメニューバーなどに紛れ込んでしまっている要素 記事のサムネイルとして表示されていて、クリック不可能な画像 	<ul style="list-style-type: none"> 下線や色による強調のないまま端にメニューとして列挙されている単語 下線や色による強調はあるが、小さい文字で密集してしまっている単語 動画投稿サイトのサムネイル画像 ニュースの記事のタイトルや動画のタイトルのようなクリック可能性を示さない文章 ブログや動画などの投稿者のプロフィールへと飛べるユーザ名 画像やイラストの中に埋め込まれている要素

図 7 に、1つの画像の中に複数のリンクが含まれている要素の例を示す。この画像のうち、各地の地名や天気を表すボックスの部分は各地の詳細な天気を確認できるページへと遷移できるリンクとなっているが、それぞれ見逃されてしまった。これらは、そのサイトの仕様を理解しなければ気付くことができない可能性のある要素であるため、強調表示によるクリック可能性の提示が有効であると考えられる。

図 8 に、クリック可能要素が密集している画面の例を示す。この画面において、太字で書かれている「〇〇から探す」という文字列のうち、緑色でハイライトを行った5つの要素はクリック可能であるが、残りの3つはクリックが不可能であり、一貫性に欠けてしまっている。この事例の場合、今回は誤検出を起こしてしまっているが、実際のユーザがサービスを利用する場面では、一度この太字部分の要素がクリックできないと判断した場合、残りのクリック可能要素もクリックできないと判断してしまう可能性がある。そのため、このように類似した要素が列挙されている箇所において誤検出が起こった場合、推測がなされていてもクリックが可能であることを示す強調表示を行う事が適切だと考えられる。

6. 操作がわかりにくいサイトでの追加検証

アクセス数上位のサイトで評価を行ったが、これらのサ

イトはアクセス数が多いこともあり、ある程度十分にデザインされていると考えられるため、ユーザが操作に困りそうなくつかのサイトを選定し、追加検証を行った。この追加検証では、シグニファイアによりクリック可能性を示しながらクリックできない要素や、シグニファイアが無かったり、多用されていたりすることによりクリック可能性が分かりづらい要素を含むという基準でサイトを選び、約 20 件のサイトを対象に実施した。



図 7: weathernews.jp の天気予報表示画面



図 6: yahoo.co.jp のニュース記事画面



図 8: hotpepper.jp の検索画面



図 9: 下の会社のアイコン 5 つのみが実際にクリックできる要素であるサイト

検証の結果、評価で用いたサイトでは見つからなかった、ユーザが操作できないと誤認するもの実際は操作できるもの、ユーザが操作できると誤認するもの実際は操作できないものがいくつか観察された。

例えば、図 9 ではページ上部の枠囲みされたアイコンはクリックできず、下部のアイコンのみがクリックが可能な要素となっている。ここでユーザは上部の枠囲みされた部分もクリックできると誤認して操作すると考えられるが、本システムも実際に上部のアイコンもクリック可能と推測しており、操作できない部分が判定可能であることが示唆された。また、「〇〇はこちら」などのように遷移を表す表現でありながらクリックできない要素があり、そうした点についても判定できることが分かった。

また、図 10 は、ページ中に様々な太字があるため、クリックできる要素であるものの、人がクリックできないと誤認してしまうものについて、VLM も認識できていなかった事例である。このように他の太文字に紛れてしまったり、クリック可能性を示すかどうか分かりづらくなってしまったりする要素は、VLM も認識できないという点は興味深く、実装した拡張機能はこのような場面において十分に機能すると考えられる。

一方、図 11 のように、人はデザイン上このインタフェース上で「会員登録へ」と「ログインへ」がクリック可能であると感じないが、VLM は「会員登録へ」と「ログインへ」をクリック可能であると判定してしまっていたものも観察された。この結果より、VLM は見た目だけではなく、「ログイン」や「会員登録」などのようにクリック可能要素となる可能性の高い言葉であれば、シグニファイアがなくてもクリック可能だと推測してしまう可能性が示唆された。そのため、人らしくクリックできないと判定させるには、プロンプトの工夫などが必要になると考えられる。

7. おわりに

本研究では、様々なサービス提供の場面において普及している Web ページの UI を対象に、ユーザが認知しづらい要素を抽出し強調表示を行うユーザ向けシステムを作成し、



図 10: 「検査済証のない建築物に係る指定確認検査機関を活用した建築基準法適合状況調査の為のガイドライン」がクリックできる要素であるサイト



図 11: 「会員登録へ」と「ログインへ」がクリックできる要素であるサイト

その精度の評価を行った。評価の結果、クリック可能な位置をクリックが可能と予測しない未認識率は約 3 割という結果となった。また、これらの精度はサイトごとに大きな差が生じてしまう可能性があることや、半数以上のサイトにおける未認識率は 1, 2 割であったことが分かった。さらに、VLM が認識しづらい要素の共通点も明らかとなった。強調表示を行う要素として適切と言える要素の抽出も行っていたが、一部の抽出された要素においては「ニュースのタイトル部分はクリックできるであろう」というような共通認識をユーザが持っていることで、クリック可能であると判断できる要素も存在した。

今後は、VLM の予測と実際のユーザの予測の間にどれ程の乖離があるかの調査を行い、より適切な認知の難しい要素の抽出を行っていく必要がある。また、今回作成した拡張機能では VLM がどの画像をクリック可能と判断しているのかを紐づけることができていなかった、今後は座標による出力や分割した画像による API の利用などによりこうした問題に対応する予定である。

一方、現状のシステムによる強調表示は単色による強調だけとなり、適切な強調方法の検討は行っていない。そのため今後はクリックが可能であることを示しつつ、サイトのデザインを損ねない表示にするため、より最適な表示を模索したり、サイトごとに対応したりするなどの工夫を行っていく予定である。さらに、Web サイトにおける操

作にはクリックによるページ遷移だけでなく、ホバーによるメニュー表示やタブによるメニューの切り替えなど様々なものがある。このような操作が必要な要素を対象に対応できる UI を増やしていく予定である。

参考文献

- [1] 相生公成: クラウド時代の IT 産業エコシステム, 産業学会研究年報, Vol. 2019, No. 34, pp. 91–111 (2019).
- [2] 広川美津雄, 井上勝雄, 岩城達也, 加島智子: 直感的インタフェースデザインの設計論の基礎的考察, 日本感性工学会論文誌, Vol. 13, No. 5, pp. 543–554 (2014).
- [3] Apple Inc.: Apple Design Tips - デザインヒント (2024). <https://developer.apple.com/jp/design/tips/> Accessed: 2024-12-20.
- [4] Google LLC: Google Design - デザインガイドラインとリソース (2024). <https://design.google/> Accessed: 2024-12-20.
- [5] Islam, M. N. and Tetard, F.: Exploring the impact of interface signs' interpretation accuracy, design, and evaluation on web usability: A semiotics perspective, *Journal of Systems and Information Technology*, Vol. 16, No. 4, pp. 250–276 (2013).
- [6] Duan, P., Warner, J., Li, Y. and Hartmann, B.: Generating Automatic Feedback on UI Mockups with Large Language Models, *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–20 (2024).
- [7] Swearngin, A. and Li, Y.: Modeling Mobile Interface Tappability Using Crowdsourcing and Deep Learning, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–11 (2019).
- [8] Țichindelean, M., Țichindelean, Monica Teodora, C. I. and Orzan, G.: A Comparative Eye Tracking Study of Usability—Towards Sustainable Web Design, *Sustainability*, Vol. 13, No. 18 (2021). [mdpi: https://doi.org/10.3390/su131810415](https://doi.org/10.3390/su131810415).
- [9] Riegler, A. and Holzmann, C.: Measuring Visual User Interface Complexity of Mobile Applications With Metrics, *Interacting with Computers*, Vol. 30, No. 3, pp. 207–223 (2018).
- [10] S. Shyam Sundar, Saraswathi Bellur, J. O. Q. X. and Jia, H.: User Experience of On-Screen Interaction Techniques: An Experimental Investigation of Clicking, Sliding, Zooming, Hovering, Dragging, and Flipping, *Human-Computer Interaction*, Vol. 29, No. 2, pp. 109–152 (2014).
- [11] 久保麻子: EC サイト/アプリにおける UX がブランド態度に与える影響, マーケティングジャーナル, Vol. 39, No. 3, pp. 32–51 (2020).
- [12] Bisante, A., Datla, V. S. V., Panizzi, E., Trasciatti, G. and Zeppieri, S.: Enhancing Interface Design with AI: An Exploratory Study on a ChatGPT-4-Based Tool for Cognitive Walkthrough Inspired Evaluations, *Proceedings of the 2024 International Conference on Advanced Visual Interfaces* (2024). [acm: https://doi.org/10.1145/3656650.3656676](https://doi.org/10.1145/3656650.3656676).
- [13] Deka, B., Huang, Z., Franzen, C., Nichols, J., Li, Y. and Kumar, R.: ZIPT: Zero-Integration Performance Testing of Mobile App Designs, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 727–736 (2017).
- [14] Schoop, E., Zhou, X., Li, G., Chen, Z., Hartmann, B. and Li, Y.: Predicting and Explaining Mobile UI Tappability with Vision Modeling and Saliency Analysis, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). [arxiv: https://doi.org/10.48550/arXiv.2204.02448](https://doi.org/10.48550/arXiv.2204.02448).
- [15] Wu, J., Zhang, X., Nichols, J. and Bigham, J. P.: Screen Parsing: Towards Reverse Engineering of UI Models from Screenshots, *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 470–483 (2021).
- [16] Wu, J., Peng, Y.-H., Li, X. Y. A., Swearngin, A., Bigham, J. P. and Nichols, J.: UIClip: A Data-driven Model for Assessing User Interface Design, *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (2024). [arxiv: https://doi.org/10.48550/arXiv.2404.12500](https://doi.org/10.48550/arXiv.2404.12500).
- [17] Gao, D., Ji, L., Bai, Z., Ouyang, M., Li, P., Mao, D., Wu, Q., Zhang, W., Wang, P., Guo, X., Wang, H., Zhou, L. and Shou, M. Z.: ASSISTGUI: Task-Oriented Desktop Graphical User Interface Automation (2024). [arxiv: https://doi.org/10.48550/arXiv.2312.13108](https://doi.org/10.48550/arXiv.2312.13108).
- [18] Ran, D., Wang, H., Song, Z., Wu, M., Cao, Y., Zhang, Y., Yang, W. and Xie, T.: Guardian: A Runtime Framework for LLM-Based UI Exploration, *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis*, pp. 958–970 (2024).
- [19] Feiz, S., Wu, J., Zhang, X., Swearngin, A., Barik, T. and Nichols, J.: Understanding Screen Relationships from Screenshots of Smartphone Applications, *Proceedings of the 27th International Conference on Intelligent User Interfaces*, pp. 447–458 (2022).
- [20] Chen, D., Huang, Y., Wu, S., Tang, J., Chen, L., Bai, Y., He, Z., Wang, C., Zhou, H., Li, Y., Zhou, T., Yu, Y., Gao, C., Zhang, Q., Gui, Y., Li, Z., Wan, Y., Zhou, P., Gao, J. and Sun, L.: GUI-WORLD: A Dataset for GUI-oriented Multimodal LLM-based Agents (2024). [arxiv: https://doi.org/10.48550/arXiv.2406.10819](https://doi.org/10.48550/arXiv.2406.10819).
- [21] Wu, J., Krosnick, R., Schoop, E., Swearngin, A., Bigham, J. P. and Nichols, J.: Never-ending Learning of User Interfaces, *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (2023). [arxiv: https://doi.org/10.48550/arXiv.2308.08726](https://doi.org/10.48550/arXiv.2308.08726).
- [22] Yuan, A. and Li, Y.: Modeling Human Visual Search Performance on Realistic Webpages Using Analytical and Deep Learning Methods, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12 (2020).
- [23] Zhou, X. and Li, Y.: Large-Scale Modeling of Mobile User Click Behaviors Using Deep Learning, *Proceedings of the 15th ACM Conference on Recommender Systems*, Association for Computing Machinery, pp. 473–483 (2021).
- [24] Bylinskii, Z., Kim, N. W., Peter O' Donovan, Alsheikh, S., Madan, S., Pfister, H., Durand, F., Russell, B., Hertzmann, A.: Learning Visual Importance for Graphic Designs and Data Visualizations, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 57–69 (2017).
- [25] 中村聡史: BADUI 診療所: カルテ 48 人なみに引かれる AI さん, ヒューマンインタフェース学会誌, Vol. 26, No. 1, pp. 22–23 (2024).
- [26] Ahrefs: Top Websites in Japan - Ahrefs (2024). Accessed: 2024-12-16.