Automated Detection of Deceptive Design Patterns (Dark Patterns) In E-Commerce Websites

Arvind Jagdish Rebari[1], Avinash Dinkar Gogawale[2], Pratiksha Rohidas Patil[3], Snehal Dattatray Dhore[4], Ashwini Sanjay Gagare[5]

[1,2,3,4,5]*Department of Science & Computer Science, MIT Arts, Commerce and Science College, Alandi (D), Pune, Maharashtra, India*

doi.org/10.64643/ijirt.208536-459

***Abstract*—Online shopping has made product discovery and purchasing faster, but the same interfaces can also contain deceptive design practices known as dark patterns. These patterns can create false urgency, hide important costs, make unwanted options more prominent, encourage forced actions, or make cancellation difficult. Such practices can affect informed consumer choice and reduce trust in digital commerce. This paper presents a proposed automated detection framework for identifying dark patterns on e-commerce websites using web scraping, user-interface feature extraction, rule-based checks, and machine learning. The system accepts a website URL, collects visible text and selected interface elements, converts them into structured features, and classifies possible patterns such as fake countdown timers, hidden costs, forced enrolment, misleading choice presentation, scarcity messages, and hard-to-cancel flows. Instead of returning only a binary decision, the proposed system stores evidence and produces a pattern label, confidence score, and human-readable explanation. The paper reviews existing research on large-scale dark-pattern detection and consumer protection, describes the proposed architecture, and discusses an evaluation framework using precision, recall, F1-score, false-positive rate, and category-level performance. The work is intended as a practical research direction for analysing e-commerce interfaces, including commonly used shopping platforms, without assuming that a particular platform contains a deceptive pattern unless the system provides evidence. The proposed approach aims to support transparent shopping, research, and responsible interface design.**

***Index Terms*—Dark Patterns, Deceptive Design, E-Commerce, Web Scraping, Machine Learning, UI Analysis, Consumer Protection, NLP, Explainable Detection**

I. INTRODUCTION

Electronic commerce depends heavily on user interfaces. Product cards, price labels, buttons, banners, pop-ups, recommendations and checkout screens are designed to help users' complete tasks. However, interface design can also be used to steer users towards choices that they may not have made if alternatives and consequences were presented clearly. Research commonly refers to such manipulative interface choices as dark patterns. Earlier work established the ethical and design concerns around dark patterns, while later large-scale measurements showed that these practices can be found across shopping websites [1], [2].

The problem is not simply that a page contains persuasive language. A normal discount message or recommendation may be legitimate. The concern begins when the interface creates a misleading impression, hides material information, preselects an option against the user's likely interest, or places unnecessary obstacles in the path to a reasonable choice. The U.S. Federal Trade Commission has highlighted practices such as hidden costs, difficult cancellation, misleading countdowns and interfaces that steer consumers towards particular choices [3]. A European Commission screening of 399 retail websites also identified fake countdown timers, choice steering and hidden information among the practices examined [4].

For an e-commerce user, identifying these patterns manually is difficult because websites are dynamic, frequently updated and visually complex. A single page can contain text, buttons, timers, pop-ups, forms, prices and links that change according to user actions. A useful detection system therefore needs to combine web collection with interface analysis rather than relying only on a fixed keyword list.

This paper proposes an automated detection framework for deceptive design patterns in e-commerce. The framework is designed around four goals: (1) collect relevant interface information from a given website, (2) identify linguistic and structural indicators of dark patterns, (3) classify the likely pattern using machine learning supported by rule-based evidence, and (4) explain the decision using the page element and text that triggered the detection. The objective is not to label a company as deceptive, but to identify interface behaviour that deserves review.

II. LITERATURE REVIEW

Dark-pattern research has developed from descriptive design studies to automated measurement and regulatory analysis. Gray et al. examined dark patterns from an ethical UX perspective and grouped examples into broader strategies such as nagging, obstruction, sneaking, interface interference and forced action [1]. Their work showed that dark patterns raise questions about the responsibility of designers when user value is replaced by business interests.

Mathur et al. presented one of the most influential large-scale studies of shopping websites [2]. Their automated approach analysed about 53,000 product pages from roughly 11,000 shopping websites and reported 1,818 dark-pattern instances across 15 types. They also identified deceptive instances across 183 websites. The study demonstrated that automated collection can support large-scale analysis, but it also noted limitations caused by dynamic pages, incomplete purchase flows and the fact that only selected pages and interface elements were examined [2].

Regulatory research has made the issue more practical. The FTC's report describes dark patterns as interface practices that can trick or manipulate consumers and discusses examples involving misleading advertisements, difficult cancellation, hidden terms and unwanted data sharing [3]. The OECD has similarly discussed dark commercial patterns as a consumer-policy issue in the digital economy [5]. The European Commission's 2023 screening found at least one of three examined manipulative practices on 148 of 399 online shops, showing that the problem is measurable at scale [4].

Recent technical research has moved towards natural language processing and transformer-based approaches. A 2025 study proposed a multi-faceted detection approach using BERT for shopping websites, showing the growing role of NLP in identifying deceptive interface language [6]. However, automated detection remains challenging because a dark pattern is often defined by the relationship between wording, visual prominence, alternatives, timing and user flow. A phrase such as 'Only 2 lefts' may be legitimate in one context and deceptive in another. This motivates a hybrid approach that combines textual, structural and behavioural evidence.

III. DARK PATTERNS IN E-COMMERCE

Dark patterns can be described as interface choices that influence users through deception, pressure, obstruction or unequal presentation of alternatives. In e-commerce, the most relevant categories for automated detection are those that leave observable evidence in page content or interaction flow.

3.1. Fake Urgency and Countdown Timers

A page may display a timer or message such as 'offer ends in 10 minutes' to create pressure. A detector should not automatically classify every timer as deceptive. Instead, it can examine whether the timer resets, whether the same deadline remains after refresh, whether the product offer is still available after the deadline, and whether the language contains urgency indicators. These signals can be combined to produce a risk score.

3.2. Hidden Costs

Hidden-cost patterns occur when a mandatory charge is not made sufficiently visible until late in the purchase process. The system can compare price information across product, cart and checkout states and detect newly introduced mandatory charges. Text extraction can also identify labels such as handling, service, platform or convenience fees.

3.3. Forced Action and Forced Enrolment

Forced action occurs when a user must complete an unrelated step to continue, such as creating an account, accepting marketing communication or sharing information when a reasonable alternative should exist. The detector can inspect forms, required fields, button labels and navigation paths to identify such behaviour.

3.4. Interface Interference and Misdirection

Interface interference occurs when visual hierarchy or wording makes one choice much easier to find than another. Examples include an oversized primary button for acceptance and a low-visibility alternative, or wording that makes one option appear socially or emotionally preferable. Such cases require a combination of DOM structure, text and visual features.

3.5. Scarcity and Social-Proof Messages

Messages such as 'only 2 lefts' or '15 people are viewing this item' may influence decisions. The proposed system treats these as signals rather than proof of deception. Repeated or inconsistent messages, especially when combined with rapidly changing values, can increase the detection score and trigger human review.

IV. PROPOSED SYSTEM

The proposed system is a web-based analysis tool in which a researcher enters an e-commerce URL. The system then crawls the relevant page or controlled purchase flow, extracts visible and structural interface information, processes the collected data, and passes the resulting features through a hybrid detection layer. The final output contains the detected pattern category, evidence, confidence and a short explanation.

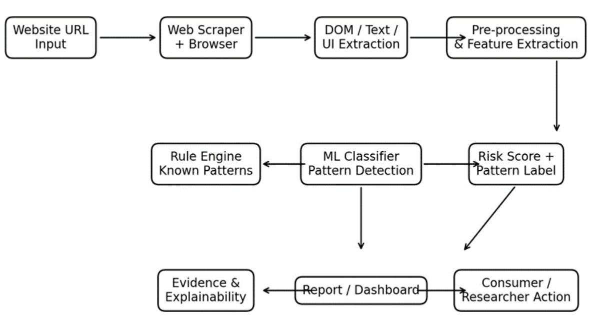

Fig. 1. Proposed architecture for automated dark-pattern detection in e-commerce.

Data Collection. The first stage uses a web scraper and, where necessary, a browser automation layer to render JavaScript-driven pages. The collector records page text, button labels, headings, links, prices, forms, timers, checkboxes, pop-ups, selected states, and navigation events. Screenshots may also be stored as evidence.

Pre-processing and Feature Extraction. Collected content is cleaned and converted into features. Text features can include urgency words, cancellation language, price-related terms, emotional pressure, opt-out wording and subscription terms. Structural features can include button prominence, number of clicks, preselected controls, hidden elements, newly appearing fees and repeated timers. These features create a representation of the interface for classification.

Rule-Based Detection. Some patterns have strong observable rules. For example, a mandatory charge that appears only at the final checkout stage can be flagged as a hidden-cost candidate. A subscription that can be started in one or two steps but requires many unrelated steps to cancel can be flagged for obstruction. Rules provide interpretable evidence and can also support the machine-learning classifier.

Machine-Learning Classification. The extracted features are passed to a classifier that predicts the likely dark-pattern category. A baseline implementation can use TF-IDF text features with a Random Forest or linear classifier, while a transformer-based model such as BERT can be evaluated for richer language understanding [6]. The proposed design does not require one fixed algorithm; the classifier can be replaced as a labelled dataset becomes larger.

Evidence and Explanation. The final stage records the page element responsible for the prediction. For example, the report may show the exact urgency text, the newly introduced fee, the preselected checkbox or the number of steps required for cancellation. This evidence makes the system more useful for researchers because the classification can be reviewed rather than treated as an unexplained model output.

V. DETECTION WORKFLOW

The complete workflow begins with URL submission and ends with a structured report. The following sequence is proposed:

- Enter Website URL: The researcher provides the target e-commerce URL and selects the page or flow to inspect.
- Render and Crawl: The system loads dynamic

content and records relevant interface states while following a controlled navigation path.

- Extract Interface Evidence: Text, buttons, links, prices, forms, timers, selected options and screenshots are collected.
- Generate Features: Linguistic, visual-structural and interaction-based features are created.
- Run Rules and ML Model: Deterministic rules identify strong signals while the classifier estimates the most likely category.
- Calculate Risk and Confidence: Evidence is combined into a category score and confidence value.
- Generate Report: The dashboard presents the pattern type, evidence, page location, confidence and a short explanation.

VI. COMPARISON WITH MANUAL AND BASIC DETECTION

Aspect	Manual / Basic Scan	Proposed Hybrid System
Scale	Limited to a few pages at a time	Can analyse many target pages or flows
Dynamic content	Depends on human observation	Browser-based rendering can capture changing UI
Text analysis	Manual reading or keywords	NLP features and classifier
Structural evidence	Hard to record consistently	DOM, controls and flow features are stored
Interpretability	Human judgement is high	Evidence and explanation are returned
Consistency	Depends on reviewer	Same detection pipeline can be repeated
False positives	Depends on human judgement	Reduced through hybrid rules, model and review

VII. DATASET AND TRAINING STRATEGY

A reliable classifier requires labelled examples. The proposed dataset should contain interface samples labelled by dark-pattern category, together with the page text and structural evidence that led to the label. Positive examples can be collected from documented dark-pattern cases and controlled examples created for research.

Negative examples should include ordinary e-commerce interfaces that use legitimate urgency, discounts, recommendations or subscription choices without deceptive behaviour.

Each record can contain the URL or anonymised page identifier, extracted text, DOM features, interaction sequence, pattern category, evidence span and reviewer label. The dataset should be divided into training, validation and test sets at the website level rather than randomly splitting individual elements from the same website. This prevents the model from learning a site's repeated design instead of learning the underlying pattern.

Because dark patterns are multi-class and sometimes overlapping, a multi-label formulation may be useful. For example, a checkout page may contain both hidden costs and interface interference. The dataset should therefore permit more than one label when reviewers agree that multiple patterns are present.

VIII. EVALUATION METHODOLOGY

The proposed system should be evaluated using a held-out test set that is not used during model development. Accuracy alone is not sufficient because some pattern categories may be less common than others. Precision measures how many detected cases are actually relevant, while recall measures how many relevant cases are found. F1-score provides a balance between the two. Evaluation should also include category-level confusion matrices, false-positive rate, false-negative rate and the quality of evidence shown to reviewers. For dynamic patterns, repeated crawling at different times can test whether the system detects changes in timers, prices or availability messages. Human review can measure whether the generated explanation is sufficient for an independent reviewer to understand why the system flagged a page.

An important research principle is that the system should not claim that a business is legally violating a rule merely because a classifier assigns a dark-pattern label. The tool is a detection and screening system. Final interpretation requires context, evidence and, where appropriate, legal or consumer-protection review.

IX. ILLUSTRATIVE E-COMMERCE CASE

Consider a product page on a large marketplace. The page displays a countdown message, a product price and a purchase button. When the researcher refreshes the page, the same countdown begins again with a similar duration. The system would not immediately call this deceptive. Instead, it would store the timer text, timestamp, page state and repeated observations. If the offer remains available after the timer expires and the timer resets, the evidence can increase the urgency-risk score.

A second example is checkout. Suppose a product is displayed at one price, but a mandatory service charge appears only near the final payment step. The system can compare the earlier and later states, identify the new mandatory charge and label the case as a hidden-cost candidate. The report can show the earlier price, the later total and the element where the charge appeared. These examples illustrate how the proposed system is intended to work on e-commerce platforms such as large fashion, general marketplace and quick-commerce sites. They are examples of the target environment, not claims that a particular platform currently uses a dark pattern.

X. BENEFITS OF THE PROPOSED APPROACH

- Scalability: automated crawling can support analysis beyond the small number of pages a human reviewer can inspect.
- Consistency: the same rules, features and model can be applied repeatedly across websites.
- Evidence-based reporting: each detection can retain the text or interface element that caused the flag.
- Early screening: researchers and consumer-protection teams can prioritise pages that require detailed review.
- Adaptability: new pattern categories can be added as interface practices evolve.
- Research support: collected evidence can be used to study how deceptive design changes across time, websites and categories.

XI. CHALLENGES AND LIMITATIONS

Dynamic websites are the first major challenge. Modern shopping pages may render important elements only after user interaction, login, location selection or asynchronous requests. A simple HTML scraper may therefore miss relevant evidence. Browser automation improves coverage but increases computation and crawling time. The second challenge is context. A countdown timer can be genuine, and a low-stock message can reflect actual inventory. Similarly, a subscription can be clearly disclosed even if it is recurring. The classifier must therefore avoid treating isolated words as proof. Combining multiple features and retaining evidence is essential.

Another limitation is dataset quality. Labels for dark patterns involve human judgement, and different reviewers may disagree about whether a design is deceptive. Annotation guidelines and multiple reviewers are needed. Class imbalance is also likely because some patterns occur much more often than others.

Finally, websites change frequently. A detector trained on old interface styles may perform poorly on new designs. Continuous evaluation and periodic retraining are therefore required. Privacy, crawling permissions, robots.txt, rate limits and responsible data collection should also be considered when building a real deployment.

XII. FUTURE SCOPE

Future versions can combine DOM analysis with computer vision so that the system understands visual hierarchy, colour contrast, button size and spatial relationships. Multimodal models could combine screenshots, page text and interaction traces in one prediction.

Another direction is temporal detection. Instead of analysing one page snapshot, the system could compare the same offer across time and detect repeated timers, changing stock messages or price transitions. A browser extension could provide real-time warnings to users, while a research dashboard could maintain a historical record of detected patterns. Explainable machine learning is also important. The system should show which features influenced the classification and which page element was responsible. This can help researchers validate the model and reduce blind reliance on automated labels. In the longer term, a standard benchmark dataset for Indian e-commerce interfaces could help researchers compare models under similar conditions.

XIII. CONCLUSION

Dark patterns create a practical challenge for trustworthy e-commerce because interface design can influence consumers without making the consequences equally clear. Existing research has demonstrated that such practices can be measured at scale, while consumer-protection agencies have documented concrete examples and risks [2]– [5].

This paper proposed a hybrid automated framework that combines web scraping, interface feature extraction, rule-based checks and machine learning. The main contribution is not simply assigning a dark-pattern label, but connecting that label to observable evidence such as wording, interface state, price changes, timers and interaction steps. This makes the system more suitable for research and review.

The proposed framework should be treated as a screening and evidence-generation tool rather than an automatic legal judgement.

Future implementation and testing should focus on high-quality labelled data, dynamic websites, multimodal analysis and transparent explanations. With careful evaluation, automated detection can support more transparent digital commerce and help researchers understand how deceptive interface practices evolve.

REFERENCES

[1] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, "The Dark (Patterns) Side of UX Design," in *Proc. 2018 CHI Conf. Human Factors in Computing Systems*, Paper 534, 2018, doi: 10.1145/3173574.3174108.

[2] A. Mathur, G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, M. Chetty, and A. Narayanan, "Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites," Proc. ACM Hum. -Comput. Interact., vol. 3, no. CSCW, Art. 81, pp. 1–32, 2019, doi: 10.1145/3359183.

[3] Federal Trade Commission, "Bringing Dark Patterns to Light: Staff Report," U.S. Federal Trade Commission, Sep. 2022.

[4] European Commission, "Consumer protection: manipulative online practices found on 148 out of 399 online shops screened," Press Release IP/23/418, Jan. 30, 2023.

[5] OECD, "Dark Commercial Patterns," OECD Digital Economy Papers, no. 336, Oct. 2022.

[6] R. Vedhapriyavadhanaa, P. Bharti, and S. Chidambaranathan, "Detecting dark patterns in shopping websites – a multi-faceted approach using Bidirectional Encoder Representations from Transformers (BERT)," Enterprise Information Systems, published online Feb. 24, 2025, doi: 10.1080/17517575.2025.2457961.

[7] H. Brignull, "Dark Patterns," darkpatterns.org, 2010.

[8] A. Mathur et al., "Dark patterns at scale: Findings from a crawl of 11K shopping websites," arXiv:1907.07032, 2019.

[9] Federal Trade Commission, "FTC Report Shows Rise in Sophisticated Dark Patterns Designed to Trick and Trap Consumers," Sep. 15, 2022.

[10] European Commission, "A safer and fairer online environment," Shaping Europe's Digital Future, 2024.

[11] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, "Dark patterns: Designing for deception," related practitioner and HCI research discussed in Proc. CHI 2018, 2018.

[12] A. Narayanan, S. Mathur, A. Chetty, and M. Kshirsagar, "Dark patterns: Past, present, and future," *Communications of the ACM*, survey and research discussion, 2020s.