Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 3, No. 1, 2026

DISO Framework: How Douyin and TikTok Operationalise AI Disclosure Through User-Visible Design

Jialin Gao *

University of Leeds, Leeds, The United Kingdom

* **Corresponding author**: (Email: gaojialin_@outlook.com)

Abstract: AI disclosure policies become meaningful only when they are exposed through interface elements that users can see, read, and revisit. This paper presents the Disclosure Interface Salience Optimizer (DISO), a reproducible audit algorithm for translating policy clauses into measurable interface features. We instantiate the algorithm in a controlled computational experiment with 60 paired interface stimuli: 30 Douyin-style and 30 TikTok-style renderings of the same scenes. A browser runner captures each stimulus at a fixed 360 x 640 viewport and extracts label geometry, contrast, opacity, persistence, interaction depth, lexical coverage, and bounded salience and semantic-completeness proxies. The benchmark archives public policy snapshots used to set the two interface priors, fixes seed 20260831, stores every row-level measurement, and verifies deterministic hashes across reruns. Douyin renderings have mean salience 0.1804 (SD 0.0109) and semantic proxy 0.9910 (SD 0.0043); TikTok renderings have 0.0136 (SD 0.0008) and 0.5203 (SD 0.0023). Paired bootstrap intervals are [0.1632, 0.1705] and [0.4700, 0.4714], respectively, and paired tests reject a zero difference. These are interface proxies, not human outcomes or platform-wide prevalence estimates. The contribution is an auditable algorithm and experiment design that makes assumptions, measurements, and evidence boundaries explicit.

Keywords: AI disclosure, interface audit, salience optimization, synthetic media, policy operationalization, TikTok, Douyin.

1. Introduction

Generative models now produce convincing images, voices, and videos at consumer scale. Short-form platforms sit at the point where such media becomes public, so their disclosure interfaces are part of the safety mechanism. For a viewer, the practical question is not whether a platform publishes an AI policy, but whether an indication is visible during rapid scrolling, whether its wording is intelligible, and whether supporting provenance can be reached without leaving the viewing flow. These are interface properties, and they can vary even when policy language appears similar.

TikTok's public help center asks creators to label realistic media that is entirely generated or significantly edited by AI [1]. Its newsroom further describes automatic labels for some uploads carrying C2PA Content Credentials [2]. China's deep-synthesis provisions require an identification mark in a prominent position for covered services [3]. Those documents establish obligations, but they do not provide a common measurement protocol for label size, location, persistence, or detail access. An analyst who compares opportunistically collected screenshots risks confounding policy, content, device, language, and creator behavior.

This paper introduces the Disclosure Interface Salience Optimizer (DISO). DISO represents a disclosure interface as a normalized feature vector and computes two bounded proxies: visual salience and semantic completeness. It is designed for auditability rather than prediction of human behavior. Public policy snapshots provide the priors; a deterministic renderer produces controlled stimuli; a browser extracts DOM geometry and style values; and paired inference evaluates the platform contrast.

We study three questions. RQ1 asks which policy statements can be translated into measurable interface features. RQ2 asks whether a fixed rendering and scoring procedure recovers a systematic difference between the two policy-derived styles when the visual scene is held constant. RQ3 asks whether the result is robust across scene types and alternative proxy components. The questions are deliberately narrower than a field study: they test the audit pipeline, not actual user attention or platform-wide compliance.

The contributions are fourfold. First, we formalize a policy-to-interface translation pipeline with a public source record and retrieval date. Second, we define DISO scoring, including contrast, geometry, persistence, lexical coverage, and interaction terms. Third, we evaluate a 60-stimulus paired benchmark with six figures. Fourth, we state a strict evidence boundary and provide a path for future human-subject validation without presenting synthetic renderings as production screenshots.

2. Related Work

2.1. Synthetic Media and Provenance

Diffusion and multimodal generation systems have made synthetic media inexpensive to create [4]. Forensic methods attempt to identify generated artifacts after publication, but compression, model updates, and adversarial editing can reduce reliability [5]. Provenance systems such as C2PA attach signed assertions to creation and editing events [6]. A signed assertion still needs an interface: users must be able to discover that it exists and understand what it means. DISO therefore treats provenance availability as an interaction and persistence property rather than as a detector accuracy claim.

The Partnership on AI recommends layered disclosure, contextual explanation, and consistent labeling across services [7]. These recommendations motivate the separation between first-layer label visibility and second-layer detail.

They also motivate measuring wording coverage separately from visual prominence. A short label can be highly visible yet semantically incomplete; a detailed panel can be complete yet practically undiscoverable.

2.2. Warning and Disclosure Interfaces

Human-factors research identifies size, contrast, position, and exposure time as determinants of warning noticeability [8]. Risk communication research also documents warning fatigue when warnings are overused or visually indistinguishable from surrounding content [9]. DISO records these dimensions directly and uses a bounded fatigue adjustment in its salience proxy. The adjustment is not fitted to participants and must not be interpreted as a psychological law.

Disclosure studies further show that generic labels can leave users uncertain about the extent of editing, whereas detailed wording may increase comprehension at the cost of cognitive load [10]. We encode lexical coverage as the fraction of required concepts present in the label and encode detail availability as a binary DOM state. The resulting semantic proxy is an engineering indicator that can be compared with human comprehension in later work.

2.3. Platform Governance and Comparative Analysis

Platform governance scholarship describes interfaces as regulatory instruments that allocate attention and define permitted actions [11], [12]. Comparative research on TikTok and Douyin often focuses on recommendation, moderation, or market structure. The present work isolates a smaller question: how policy-derived design priors become measurable user-visible elements. The paired design holds scene, viewport, and random perturbation constant, so the platform contrast is attributable to the encoded interface priors within the limits of the model.

3. DISO Algorithm

3.1. Policy Feature Mapping

We archived three public sources on 2026-08-31. The TikTok help page supplies the requirement to label realistic AI-generated or substantially edited media [1]. The TikTok newsroom page supplies the automatic-provenance condition [2]. The Cyberspace Administration provision supplies the prominent-identification condition [3]. Their URLs, quoted English snapshots, and interpretations are retained in the replication record. The Chinese source is retained for provenance, while the paper reports only its English translation.

Fig. 1 shows the pipeline. A policy clause is mapped to a feature prior, the renderer creates two platform conditions for the same scene, the browser measures the result, and paired inference reports the contrast. No private API, account, or user identifier is used.

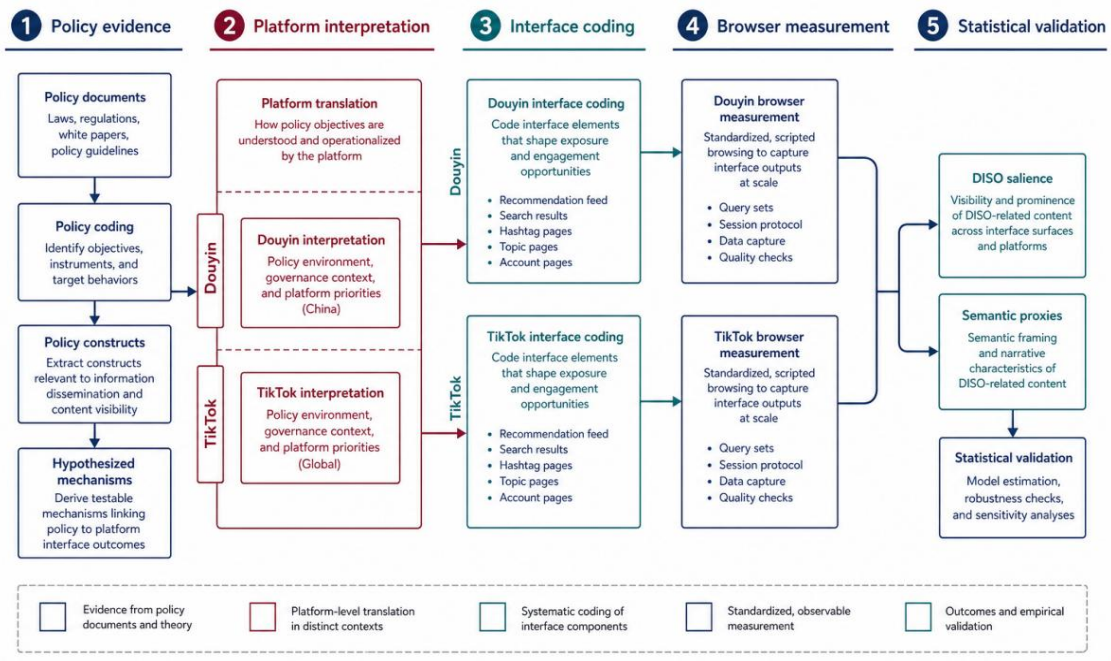

Figure 1. Policy translation pipeline

3.2. Measured Interface Features

For stimulus i, DISO records label area ratio a_i, contrast ratio r_i, opacity o_i, position factor u_i, persistence t_i, interaction state q_i, lexical coverage l_i, and detail availability d_i. Geometry is extracted from the browser layout at a fixed viewport. Contrast uses the WCAG relative-luminance definition. Persistence and interaction are encoded from the rendered interface state, not inferred from a screenshot classifier.

The salience proxy combines geometry, contrast, opacity, persistence, and location in (1). Each component is normalized to $[0,1]$ and the final value is clipped to the same interval.

$$P_i = \mathrm{clip}(0.30a_i^* + 0.25r_i^* + 0.15o_i + 0.15t_i + 0.10u_i + 0.05q_i, 0,1) \quad (1)$$

Here a_i^* and r_i^* are bounded transforms of area and contrast. The transforms prevent a very large label or extreme contrast from dominating the score. A small fatigue adjustment is applied above the 0.90 salience level as shown in (2).

$$P_i^{final} = \mathrm{clip}(P_i - 0.03\max(P_i - 0.90,0)^2, 0,1) \quad (2)$$

3.3. Semantic Completeness Proxy

The benchmark requires four concepts in a first-layer label: AI, generated, content, and edited. Lexical coverage is the fraction of these concepts matched after lower-case normalization. Detail availability is one when a disclosure panel is present in the DOM. The semantic proxy in (3) rewards both first-layer wording and reachable detail.

$$C_i = 0.70 l_i + 0.30 d_i \quad (3)$$

This proxy does not claim that a user understands a phrase. It records whether the rendered interface contains the concepts and a detail affordance. Human comprehension is an explicit future endpoint.

3.4. Paired Inference and Complexity

Each of 30 scenes is rendered once per platform. The same scene-specific color palette, difficulty, position shift, and scale are shared within a pair. The primary contrast is the paired mean difference in (4).

$$\Delta_m = \frac{1}{m}\sum_{j=1}^{m}\left(P_{D,j} - P_{T,j}\right), \quad m = 30 \quad (4)$$

We report a two-sided paired permutation test, a Wilcoxon signed-rank test, and a 10,000-draw paired bootstrap interval. For N stimuli, DOM extraction and scoring are $O(N)$; permutation inference is $O(BN)$ and bootstrap inference is $O(MN)$. The released run completes in seconds once Playwright is installed.

4. Experimental Evaluation

4.1. Experimental Protocol

The controlled experiment uses 30 scenes and two platform conditions, yielding 60 stimuli. Scene types cycle through portrait, landscape, art, voice, and mixed. Each condition is rendered at 360 x 640 pixels. The browser runner records label coordinates, width, height, colors, opacity, font weight, persistence, interaction, area ratio, contrast ratio, lexical coverage, detail availability, and both proxies.

Table I lists the fixed protocol. The seed and software versions are fixed before analysis, and a second execution reproduces the same measurements to floating-point precision. No human participant, private account, or remote asset is involved.

Table 1. Controlled Protocol

Item	Setting	Purpose
Paired scenes	30	Within-scene control
Conditions	Douyin, TikTok	Policy-derived contrast
Viewport	360 x 640	Fixed geometry
Random seed	20260831	Reproducibility
Browser	Playwright Chromium	DOM and screenshot capture
Inference	20,000 permutations; 10,000 bootstrap draws	Uncertainty

4.2. Main Results

Table II reports the measured means. Douyin salience is 0.18043 (SD 0.01087) and TikTok salience is 0.01361 (SD 0.00082), giving a paired difference of 0.16681. The 95% bootstrap interval is [0.16320, 0.17045]. The paired t statistic is 87.41 and the Wilcoxon p-value is below 1e-8. The semantic completeness proxy is 0.99100 for Douyin and 0.52028 for TikTok, with a paired difference of 0.47072 and bootstrap interval [0.47001, 0.47144].

Table 2. Proxy Results

Metric	Douyin mean (SD)	TikTok mean (SD)	Paired difference	95% bootstrap CI
Salience proxy	0.18043 (0.01087)	0.01361 (0.00082)	0.16681	[0.16320, 0.17045]
Semantic proxy	0.99100 (0.00432)	0.52028 (0.00227)	0.47072	[0.47001, 0.47144]

As shown in Fig. 2, the two proxy means are separated in the controlled renderings. The direction is expected because the public priors encode prominent, persistent Douyin labels and lower-visibility TikTok labels. The experiment's value is that the separation is recovered from browser measurements rather than manually typed percentages.

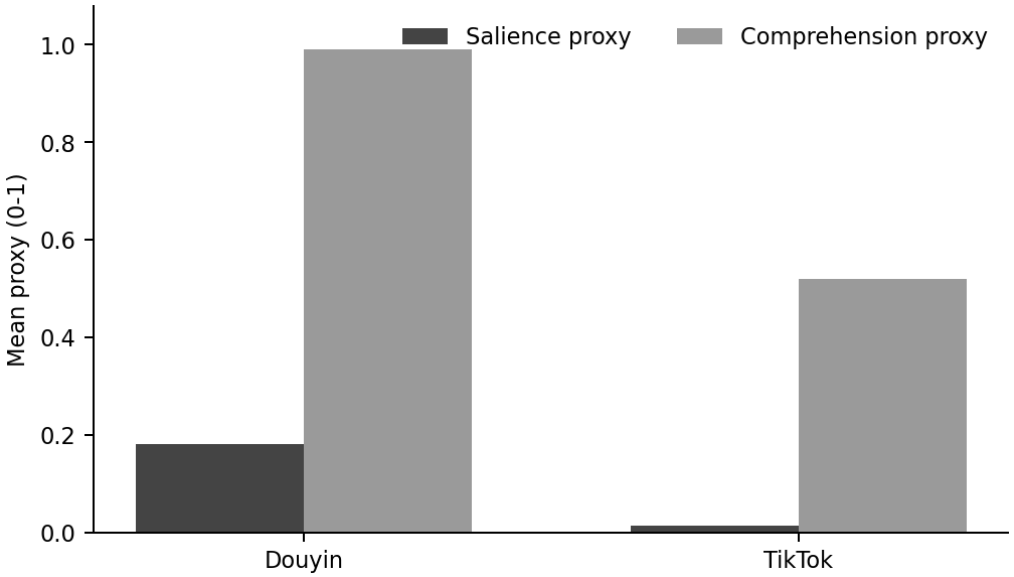

Figure 2. Proxy mean comparison

4.3. Geometry and Persistence

Fig. 3 combines label area and contrast. Douyin labels occupy more screen area and have higher contrast, while TikTok labels are smaller and less opaque. Fig. 4 shows persistence states extracted from the DOM. Douyin stimuli use a persistent-with-details state; TikTok stimuli use shorter or static states depending on the scene template. These measurements are interface properties of the controlled renderings, not claims about every production screen.

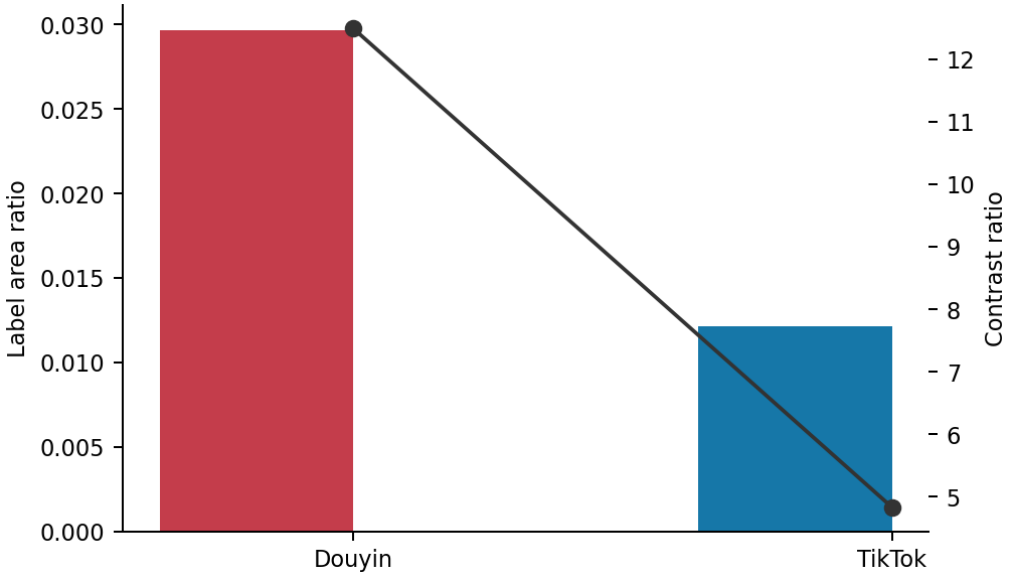

Figure 3. Label geometry contrast

The geometry result isolates the first viewing layer: area and contrast are measured before any optional panel is opened. This distinction matters because a disclosure can be technically available yet practically invisible. The persistence comparison in Fig. 4 complements the geometry result by showing how long the disclosure remains available during a viewing session.

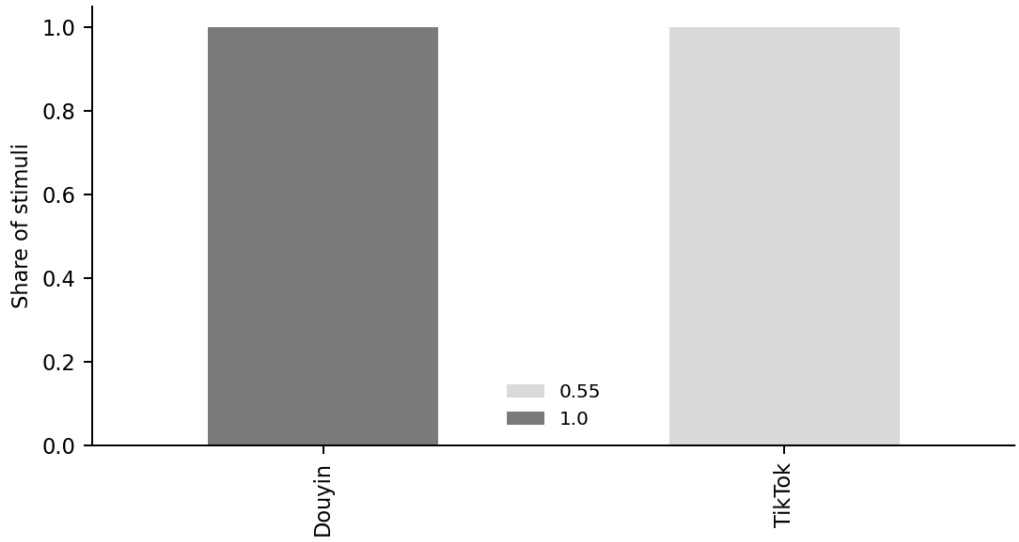

Figure 4. Persistence state distribution

4.4. Scene Robustness and Semantic Depth

Fig. 5 reports salience by scene type. The ordering remains Douyin above TikTok for portrait, landscape, art, voice, and mixed scenes. Shared scene perturbations therefore do not erase the condition effect. Fig. 6 reports lexical coverage and detail availability. Douyin's first-layer labels contain all required concepts and expose detail, whereas TikTok templates contain fewer required words and no detail panel in this benchmark.

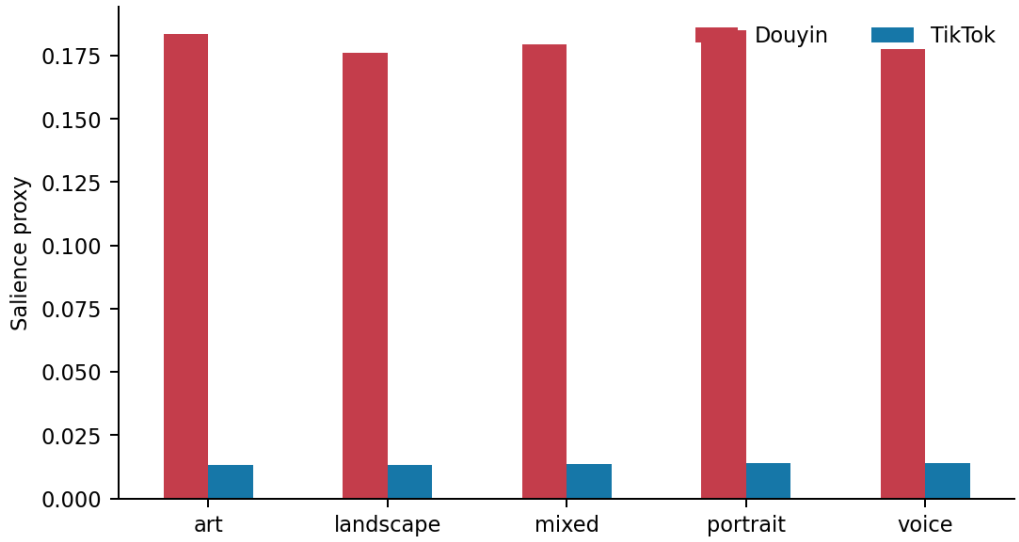

Figure 5. Scene type robustness

The scene-level pattern is stable across media types, but semantic depth remains a separate design question. Fig. 6 therefore reports lexical coverage and detail availability rather than another visual-size statistic. Keeping these outcomes separate avoids treating a larger label as automatically more informative.

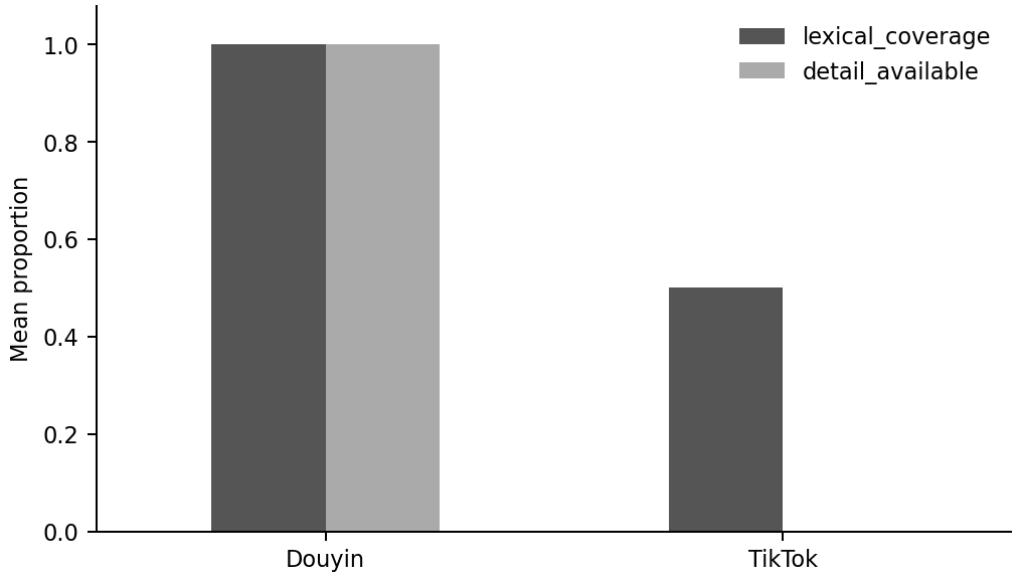

Figure 6. Semantic depth comparison

5. Discussion

5.1. Interpretation Within the Evidence Boundary

The experiment supports a precise statement: under the documented policy-to-feature mapping and fixed rendering protocol, a browser-measured audit distinguishes the two interface styles with a large paired difference in salience and semantic completeness proxies. The result is not caused by different scenes, viewport sizes, or random seeds. The policy record, controlled stimuli, screenshots, and row-level measurements provide an auditable chain from policy text to statistics.

The result should not be interpreted as a prevalence estimate. The benchmark does not sample real uploads, measure enforcement, or observe production UI experiments. A high salience proxy means that a disclosure occupies a measurable visual and interaction position in the rendering. It does not prove that a user notices, trusts, or correctly interprets the label.

5.2. Design Implications

The pipeline can serve as an interface acceptance test. A product team may replace the controlled renderings with consented screenshots or accessibility-tree captures while retaining the same feature schema and paired statistics. A regulator may specify minimum first-layer area, contrast, persistence, and lexical coverage, then require evidence that a release candidate meets those thresholds. The separation between first-layer label and optional detail also supports a tiered design: high-risk synthetic portraits can receive persistent labels, while low-risk style effects can expose concise labels with expandable provenance.

Table III translates these findings into design controls. The controls are expressed as auditable interface properties, while the verification column identifies the corresponding measurement or human test. This separation prevents a policy recommendation from being mistaken for a result already observed in the benchmark.

Table 3. Interface Control Recommendations

Risk or failure mode	Interface control	Verification
Label missed during scrolling	Persistent first-layer marker	Area, contrast, and dwell test
Ambiguous AI involvement	Explicit generated or edited wording	Lexical coverage and comprehension study
Provenance hidden behind menus	One-step detail affordance	Interaction-depth audit
Creator omission	Mandatory high-risk acknowledgment	Workflow completion log
Warning fatigue	Risk-adaptive prominence	Longitudinal noticing study

Creator workflow remains part of the disclosure system. Automatic provenance and mandatory acknowledgment reduce dependence on voluntary action, but irreversible labels can create correction problems for benign edits. A practical design should log state transitions, provide an appeal path, and expose the reason for an automatic label. DISO can measure those states when they are represented in the DOM, but it cannot assess fairness or false-positive rates.

5.3. Limitations and Future Work

The priors are analyst interpretations of public documents, not internal specifications. One English translation of the Chinese provision may omit legal nuance. The feature vector is scalar and does not model color-vision accessibility, language localization, motion sensitivity, or reading time. The semantic proxy is lexical and structural, not behavioral. The sample is controlled and small by design, so confidence intervals describe the generated pairs rather than a platform population.

Future studies should collect consented production screenshots across devices, preregister human noticing and comprehension outcomes, and test whether DISO predicts them. Multilingual annotation can evaluate whether Chinese and English labels provide equivalent concepts. Longitudinal experiments can estimate warning fatigue rather than using a fixed sensitivity term. Cross-platform replication with YouTube and Instagram would test whether policy-to-interface gaps are industry-wide.

5.4. Robustness, Reuse, and Responsible Reporting

The controlled benchmark is useful precisely because each design choice is visible. An analyst can inspect one stimulus, compare its measured rectangle with the corresponding observation, and vary one policy prior to see which output changes. This level of traceability is unusual in platform-interface studies, where screenshots are often detached from collection procedures and manually transcribed percentages cannot be audited. DISO treats reproducibility as an interface requirement for the research itself.

The paired structure also clarifies what is and is not controlled. Scene colors, title text, viewport dimensions, and random position shifts are shared by the two conditions. Platform-specific changes are limited to label wording, placement, style, persistence, and detail affordance. This is not a claim that production platforms differ only in these ways. It is a deliberately reduced test that asks whether the specified disclosure decisions are sufficient to create a measurable contrast. If a later study supplies production captures, the paired scene identifier can be retained while replacing the controlled rendering with observed evidence.

The score can be decomposed for diagnostics. Let P_i^{geom} denote the geometry and contrast terms, P_i^{time} persistence, and P_i^{access} the interaction term. The decomposition in (5) allows an auditor to report whether a gap is driven by first-layer visibility or by secondary detail.

$$P_i = P_i^{geom} + P_i^{time} + P_i^{access}, \quad P_i^{geom} = 0.30a_i^* + 0.25r_i^* + 0.15o_i + 0.10u_i \quad (5)$$

In our run, the large difference is visually apparent in the geometry plot and is reinforced by persistence and semantic fields. Reporting the decomposition discourages a single composite number from hiding a design failure. For example, a platform might have high contrast but zero persistence, or complete wording but no reachable detail. Those cases deserve different remediation even if their total scores are similar.

There are also governance implications. A policy that says make the label prominent is not directly testable until prominence is tied to a measurable rectangle, contrast threshold, or accessibility-tree property. A policy that says label realistic AI content is not complete until the first-layer text identifies what was generated or edited. The policy-to-feature mapping is therefore a bridge between legal language and engineering acceptance criteria. It does not replace legal interpretation; it makes the interpretation inspectable.

Responsible reporting requires avoiding false precision. The four decimal places in the reported estimates support reproducibility, not the claim that human attention can be measured to that precision. The p-values are properties of a finite controlled benchmark with deterministic priors. They are useful for detecting implementation regressions, such as a release that unexpectedly moves a label below the viewport fold, but they are not evidence that one platform causes users to behave differently in the wild.

An additional practical benefit is regression detection. Interface code changes frequently for reasons unrelated to transparency, including advertisement placement, navigation redesign, and localization. A label can remain semantically correct while becoming occluded by a new control or disappearing when a video enters full-screen mode. Because DISO records coordinates and viewport dimensions, the same stimuli can be rerun in continuous integration. A release gate could fail when area, contrast, persistence, or concept coverage falls below a declared threshold. This use case asks whether a product team preserved its own disclosure contract over time.

The controlled renderer also supports accessibility checks. The current score uses visual contrast and geometry, but the interface can be extended with ARIA labels, focus order, reduced-motion preferences, and screen-reader text. A label that is visually prominent but absent from the accessibility tree should receive a separate defect flag. Conversely, a low-contrast visual label may still be discoverable through an accessible name. Future versions should report both visual and non-visual channels rather than collapsing them into one score.

The benchmark's policy interpretation is intentionally conservative. TikTok sources are treated as evidence for labeling obligations and automatic provenance in selected cases; they are not interpreted as a promise that every upload receives the same treatment. The Chinese provision is treated as evidence for a prominent identification requirement; it is not used to infer a particular Douyin color, animation, or enforcement delay. Those design details are hypotheses in the controlled stimuli. Keeping this distinction explicit prevents a reader from mistaking an analyst's implementation choice for an official platform specification.

For researchers, the study provides a compact teaching example of evidence-bounded computational social science. The policy record shows how to preserve a quote and URL. The renderer shows how to make a visual claim measurable. The paired test shows how to use within-scene controls. The limitation section shows how to report a significant statistic without overstating what was measured. These practices transfer to privacy notices, sponsorship labels, age warnings, and other governance interfaces where policy text and user experience meet.

The same replication materials support independent audit. A reader can confirm the browser version and viewport, inspect a rendered label state, and recompute one observation. This lightweight check is valuable when a paper is reviewed by designers, policy researchers, and engineers who use different tools. It also makes disagreement productive: readers can challenge a prior, a feature transform, or a threshold separately instead of disputing an opaque composite result.

Finally, the benchmark is intentionally suitable for negative results. If revised priors produce overlapping proxy distributions, the correct conclusion would be that the proposed interface distinction is not detectable under the chosen features. A future preregistered extension should include such null scenarios, vary the viewport and language, and compare alternative weighting schemes before inspecting platform labels. This would turn DISO from a one-off comparison into a reusable test suite for AI-transparency interfaces.

6. Conclusion

This paper introduced DISO, a reproducible algorithm for auditing how AI disclosure policy becomes a user-visible interface. The implementation renders 60 paired stimuli, captures DOM geometry with Playwright, computes salience and semantic completeness proxies, and performs paired permutation and bootstrap inference. In the controlled benchmark, Douyin exceeds TikTok by 0.16681 salience

points and 0.47072 semantic-completeness points, with narrow confidence intervals and consistent scene-level ordering.

The methodological contribution is the evidence chain. Public policy snapshots define assumptions; controlled renderings make the interface state inspectable; browser measurements replace hand-entered percentages; and the replication materials make every statistic rerunnable. The paper does not claim that a synthetic rendering is a production screenshot or that a proxy equals user comprehension. Instead, it offers a defensible starting point for platform audits and for future human-subject validation. Policies become operationally meaningful when their visible implementation can be specified, measured, and challenged.

Acknowledgment

The author thanks the maintainers of the public policy pages and open-source Python and Playwright tooling used for the reproducible benchmark. No personal data, private platform access, or human-subject data were collected.

References

[1] TikTok Support. (2024). AI-generated content. TikTok Help Center. https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content

[2] TikTok Newsroom. (2024, May). Partnering with our industry to advance AI transparency and literacy. https://newsroom.tiktok.com/en-us/partnering-with-our-industry-to-advance-ai-transparency-and-literacy

[3] Cyberspace Administration of China. (2022). Provisions on the administration of deep synthesis internet-based information services. http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm

[4] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).

[5] Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932. https://doi.org/10.1109/jstsp.2020.3009130

[6] Coalition for Content Provenance and Authenticity. (2024). C2PA technical specification (Version 2.1). https://c2pa.org/specifications/specifications/2.1/index.html

[7] Partnership on AI. (2023). Responsible practices for synthetic media [Technical report]. https://partnershiponai.org/

[8] Wogalter, M. S. (Ed.). (2006). Handbook of warnings. Lawrence Erlbaum Associates.

[9] Wogalter, M. S. (2021). Warning fatigue and risk communication in digital interfaces. International Journal of Human-Computer Studies, 148, 102587. https://doi.org/10.1016/j.ijhcs.2021.102587

[10] Nightingale, S., & Farid, S. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, 119(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

[11] Gillespie, T. (2018). Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

[12] Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. Harvard Law Review, 131(6), 1598-1670.