

Proposition de sujet de P-SAT — Année 2025-2026

Analyse de la personnalisation et des dark patterns via la plateforme crowdsourcing STETOSCOPE

Encadrant.e.s au département informatique

Antoine BOUTET (antoine.boutet@insa-lyon.fr)

Résumé

Contexte scientifique

La personnalisation des contenus Web et mobiles ainsi que les « dark patterns » sont désormais omniprésents dans l'écosystème numérique, influençant les choix des utilisateurs de manière opaque et parfois manipulatrice. Ces pratiques se manifestent sous différentes formes : prix variables selon le profil utilisateur, classements de recherche biaisés, ou encore compteurs de stock trompeurs exploitant les biais cognitifs humains.

L'analyse de ces phénomènes à grande échelle se heurte toutefois à des obstacles techniques majeurs. Les approches traditionnelles basées sur des robots d'exploration (bots) sont limitées par les systèmes anti-fraude des plateformes qui détectent et bloquent ces agents automatisés. STETOSCOPE propose une solution innovante à ce défi : une plateforme de crowdsourcing qui collecte des données authentiques (captures d'écran) auprès d'utilisateurs réels, permettant ainsi d'auditer la transparence des algorithmes sur le Web et le mobile sans déclencher les défenses anti-bots.

Objectifs

Ce projet P-SAT vise à auditer les pratiques de manipulation de plusieurs plateformes majeures du commerce en ligne (Amazon, Booking, AliExpress, Temu). Plus spécifiquement, nous cherchons à détecter et quantifier quatre types de manipulations : la discrimination par les prix (personnalisation tarifaire), le tri biaisé des résultats de recherche, la véracité des compteurs d'urgence (stocks limités, nombre de vues), et l'authenticité des rabais promotionnels.

Pour atteindre ces objectifs, le projet s'articule autour de deux axes techniques : d'une part, faire évoluer STETOSCOPE vers un outil véritablement multi-plateforme (Android, iOS, Web) pour élargir la base de participants ; d'autre part, automatiser l'extraction de données complexes à partir des captures d'écran grâce à des techniques avancées (LLM, OCR, vision par ordinateur).

Méthodologie

La démarche adoptée repose sur cinq piliers complémentaires :

- **Conception de protocoles d'audit** : Définition rigoureuse de scénarios d'étude reproductibles via l'interface d'administration de STETOSCOPE, garantissant la cohérence des données collectées.
- **Extraction intelligente par LLM** : Développement de scripts Python exploitant des modèles de vision (type GPT-4 Vision) pour extraire automatiquement des données structurées (JSON/CSV) à partir des captures d'écran, dépassant ainsi les limites des expressions régulières traditionnelles.
- **Visualisation et analyse avancées** : Utilisation de Power BI pour transformer les données brutes en visualisations interactives facilitant l'identification de patterns discriminatoires et leur communication auprès d'un public non-technique.

- **Extension multi-plateforme** : Développement d'une interface Web complémentaire permettant l'intégration d'utilisateurs iOS et Desktop, jusqu'ici exclus de l'application Android native.
- **Campagnes participatives** : Organisation et coordination de campagnes de collecte auprès d'une population diverse de participants volontaires, assurant la représentativité et l'authenticité des données.

Mots-clés

Dark Patterns, Personnalisation des prix, LLM, Power BI, Crowdsourcing.

Contexte de travail

Ce projet s'inscrit dans le cadre du cours P-SAT (Projet Scientifique, Artistique et Technique) sous la direction d'Antoine Boutet, enseignant-chercheur au département Informatique de l'INSA Lyon. L'équipe projet est composée de trois étudiants de 5^e année du département IF.

Collaboration interdisciplinaire : Le projet bénéficie d'une collaboration enrichissante avec le département Télécommunications, Services et Usages (TC). Isabelle Ott Kiraly (isabelle.ott-kiraly@insa-lyon.fr) et Andy Vu Ngoc (andy.vu-ngoc@insa-lyon.fr) contribuent activement à la rédaction des protocoles d'étude sur STETOSCOPE, à l'organisation des campagnes de collecte de données, ainsi qu'à l'analyse bibliographique et à l'état de l'art.

Support technique et ressources : Le développement s'appuie sur les travaux initiaux de Tao Beaufils (tao.beaufils@inria.fr), développeur principal du logiciel STETOSCOPE lors de son Projet de Fin d'Études (PFE). Des échanges réguliers avec lui permettent de mieux comprendre l'architecture logicielle existante et de planifier les évolutions futures. Pour les phases de développement et de test, deux smartphones Android sont mis à disposition de l'équipe. Le suivi du projet est assuré par des réunions hebdomadaires avec l'encadrant, garantissant un accompagnement continu et une réactivité face aux difficultés rencontrées.

1 Introduction

1.1 Contexte

L'économie numérique contemporaine s'appuie massivement sur le profilage des utilisateurs et la personnalisation algorithmique des contenus. En théorie, cette personnalisation vise à améliorer l'expérience utilisateur en proposant des produits, services ou informations adaptés aux préférences individuelles. Dans la pratique, cependant, ces systèmes de recommandation fonctionnent comme des « boîtes noires » algorithmiques : ni les utilisateurs ni les régulateurs ne savent précisément quelles données personnelles alimentent ces profils, ni selon quels critères les décisions d'affichage sont prises.

Ce manque de transparence crée un terrain propice aux pratiques abusives. On observe ainsi des phénomènes de discrimination par les prix (un même produit affiché à des tarifs différents selon le profil utilisateur) ou de discrimination par la recherche (des résultats ordonnés différemment pour orienter les choix d'achat). S'ajoutent à cela les « dark patterns », ces mécanismes d'interface délibérément conçus pour exploiter les biais cognitifs humains : faux compteurs de stock créant un sentiment d'urgence, promotions fallacieuses, ou encore boutons trompeurs compliquant la désinscription.

1.2 Problématique

Auditer ces pratiques à grande échelle représente un défi technique et méthodologique considérable. Les approches traditionnelles d'audit algorithmique reposent sur des robots d'exploration (bots) qui simulent le comportement d'utilisateurs en parcourant systématiquement les plateformes Web. Cette méthode présente toutefois trois limitations fondamentales :

1. **Absence d'historique authentique** : Les bots ne possèdent pas l'historique de navigation, les cookies ou le profil comportemental d'un utilisateur réel. Or, c'est précisément sur ces données que

repose la personnalisation. Un bot voit donc une version « neutre » ou « par défaut » de la plateforme, pas la version personnalisée que verrait un utilisateur lambda.

2. **Détection et blocage** : Les plateformes commerciales ont développé des systèmes anti-fraude sophistiqués (CAPTCHA, fingerprinting, analyse comportementale) qui détectent et bloquent rapidement le trafic automatisé, rendant les campagnes d'audit à grande échelle techniquement impossibles sans contournement constant.
3. **Difficultés sur mobile** : L'automatisation sur smartphones (via des frameworks d'instrumentation comme Frida ou Appium) nécessite souvent un accès root ou des modifications système détectables par les applications. De plus, l'écosystème mobile (iOS notamment) est particulièrement fermé, rendant l'automatisation encore plus complexe que sur le Web.

Ces obstacles expliquent pourquoi, malgré l'importance sociétale des dark patterns, les études empiriques à grande échelle restent rares, particulièrement sur les plateformes mobiles qui représentent pourtant la majorité du trafic e-commerce.

1.3 Objectifs du projet

L'objectif principal de ce P-SAT est de contourner ces limitations méthodologiques en adoptant une approche radicalement différente : le crowdsourcing, ou participation citoyenne volontaire. Plutôt que de chercher à simuler artificiellement des utilisateurs réels, nous mobilisons de véritables utilisateurs humains pour collecter des données authentiques sur leur expérience personnalisée des plateformes.

Concrètement, le projet s'appuie sur STETOSCOPE, une plateforme de collecte collaborative, pour atteindre trois objectifs opérationnels :

- **Auditer les pratiques de personnalisation** sur des plateformes commerciales majeures (Amazon, Booking.com, AliExpress, Temu) en collectant les données d'affichage réelles vues par différents utilisateurs pour un même produit ou une même recherche.
- **Détecter et documenter des manipulations concrètes** : variations de prix injustifiées (discrimination tarifaire), différences dans l'ordre des résultats de recherche (discrimination algorithmique), faux compteurs de stock ou de vues (dark patterns d'urgence), promotions trompeuses (réductions fictives).
- **Automatiser l'extraction et l'analyse de données** visuelles complexes à partir des captures d'écran fournies par les participants, en exploitant des technologies avancées : reconnaissance optique de caractères (OCR), expressions régulières (Regex), et modèles de vision par apprentissage profond (LLM multimodaux).

Ce projet s'inscrit ainsi dans une démarche de transparence algorithmique et de protection des consommateurs, en fournissant des preuves empiriques des pratiques de manipulation numérique.

2 État de l'art

2.1 La collecte de données et la personnalisation

Le profilage est la norme sur Internet. Des études récentes montrent que même les assistants vocaux profilent les utilisateurs pour cibler les publicités. Pourtant, déterminer précisément quelles informations sont collectées reste un défi. Des discriminations de prix (proposer un prix différent pour le même produit selon l'utilisateur) et de recherche (ordonner les résultats différemment pour influencer l'achat) ont été documentées, mais leurs causes racines restent mal comprises.

2.2 Les outils de mesure existants

Les outils actuels pour détecter les manipulations se classent en plusieurs catégories :

- **Systèmes basés sur des règles** : Ils scannent le code des pages web à la recherche de signaux prédéfinis (ex : bannières de cookies trompeuses). Cette méthode échoue face aux tactiques subtiles ou changeantes.

- **Systèmes basés sur les données (Machine Learning) :** Ils utilisent des modèles pour identifier visuellement ou textuellement des dark patterns.
- **Approches hybrides :** Outils comme UI Guard qui analysent à la fois la mise en page et le langage.

Cependant, avec la migration massive des usages vers le mobile, l'automatisation de ces audits est devenue ardue, les plateformes utilisant des techniques de « fingerprinting » avancées pour exclure tout trafic non-humain. C'est ici que l'approche participative de STETOSCOPE se distingue, en réintégrant l'humain dans la boucle de collecte.

3 Solution Proposée : STETOSCOPE

3.1 Vue d'ensemble de l'architecture

STETOSCOPE est une plateforme composée d'une application mobile (pour les participants) et d'un serveur backend (pour l'administration et l'analyse). Le fonctionnement repose sur quatre étapes clés :

1. **Configuration de la campagne :** L'administrateur définit un scénario, les URL cibles et les instructions de guidage.
2. **Guidage de l'utilisateur :** Le participant sélectionne une campagne dans l'application. Une bannière en surimpression (overlay) le guide pas à pas dans l'application tierce (ex : Amazon) jusqu'à l'information cible.
3. **Capture :** L'utilisateur déclenche manuellement une capture d'écran. Celle-ci est envoyée au serveur avec des métadonnées (modèle du téléphone, localisation si autorisée).
4. **Analyse automatique :** Le serveur traite les captures pour extraire les données pertinentes (prix, rang des produits) et générer des visualisations.

3.2 Implémentation Technique

- **Application Mobile :** Développée en Java pour Android. Elle nécessite des permissions spécifiques pour s'afficher « par-dessus » les autres applications (overlay) afin de guider l'utilisateur sans quitter le contexte de navigation.
- **Backend et Dashboard :** L'interface d'administration est construite avec le framework Vite (React/TypeScript) pour le frontend et FastAPI pour l'API REST.
- **Traitement des données :** L'outil permet d'extraire le texte des images via OCR et d'appliquer des expressions régulières (Regex) pour structurer les données. Dans le cadre de ce P-SAT, nous explorons également l'utilisation de LLM (Large Language Models) pour améliorer cette extraction sur des données complexes.

4 Campagnes de Collecte et Résultats

Nous avons mené des campagnes préliminaires entre novembre et décembre 2025 sur plusieurs plateformes (AliExpress, Amazon, Booking, Temu, etc.). Voici les principaux résultats obtenus.

4.1 Discrimination par les prix (AliExpress)

L'objectif était de vérifier si le prix d'un même article variait selon l'utilisateur.

- **Résultat :** AliExpress est la plateforme où ce comportement est le plus flagrant. Pour un même modèle d'écouteurs, les prix affichés variaient entre 3€ et 8,50€ selon les utilisateurs.
- **Analyse :** Aucune corrélation n'a été trouvée avec le modèle de smartphone utilisé, suggérant une personnalisation basée sur le profil utilisateur ou l'historique.

4.2 Discrimination de recherche (Amazon)

Cette campagne visait à analyser l'ordre des résultats pour une recherche identique (ex : « ordinateur portable »).

- **Résultat** : Une grande disparité a été observée. Le prix moyen des 3 premiers résultats varie considérablement d'un utilisateur à l'autre, allant de 235€ à plus de 1100€.
- **Interprétation** : Cela indique que la plateforme oriente certains utilisateurs vers des gammes de produits beaucoup plus chères dès les premiers résultats.

4.3 Fausses promotions (Temu & Black Friday)

Nous avons analysé l'évolution des prix et des promotions affichées avant, pendant et après le Black Friday.

- **Résultat** : Des pratiques abusives ont été détectées sur Temu. Un article affiché à 136€ avec une réduction de « -72% » pendant le Black Friday a été retrouvé à 150€ après l'événement.
- **Constat** : La réduction affichée (-72%) était totalement décorrélée de la variation réelle du prix, constituant potentiellement une pratique commerciale trompeuse utilisant des dark patterns (urgence, couleurs vives).

4.4 Compteurs incohérents (Booking)

L'analyse des messages incitatifs (« Plus que 2 chambres disponibles », « 10 personnes regardent cet hôtel ») a révélé des incohérences.

- **Résultat** : Sur Booking.com, un compteur d'expérience a montré une valeur décroissante au fil du temps, ce qui est illogique pour un compteur cumulatif, suggérant soit un bug, soit un compteur factice.

5 Discussion

5.1 Apports du projet

STETOSCOPE démontre qu'une approche participative est viable et nécessaire pour auditer les algorithmes de personnalisation mobile. En impliquant l'utilisateur, nous contournons les défenses anti-bots et collectons des données authentiques (« User-centric »). L'outil a permis de mettre en évidence des preuves tangibles de discriminations tarifaires et de fausses promotions.

5.2 Limitations et Éthique

L'approche présente des défis. Bien que nous puissions observer le résultat de la personnalisation (l'output), le système reste une boîte noire quant aux critères exacts utilisés (l'input). De plus, la collecte de captures d'écran, bien que floutées pour les parties non pertinentes, soulève des enjeux de confidentialité qui ont été validés par le DPO (Délégué à la Protection des Données).

6 Perspectives

Conformément aux objectifs du P-SAT, les développements futurs se concentrent sur :

1. **L'extension multi-plateforme** : Portage de STETOSCOPE sur iOS et Web pour toucher une population plus large.
2. **L'analyse avancée par LLM** : Remplacement des Regex rigides par des modèles de vision (type GPT-4 Vision) pour extraire des données non structurées plus efficacement.
3. **La visualisation** : Intégration de tableaux de bord Power BI pour faciliter l'interprétation des résultats par des non-techniciens.

7 Conclusion

Ce projet P-SAT a permis de démontrer la viabilité et l'efficacité de l'approche crowdsourcing pour l'audit des algorithmes de personnalisation et la détection de dark patterns. STETOSCOPE offre une alternative crédible aux méthodes automatisées traditionnelles en intégrant l'utilisateur réel dans le processus de collecte de données.

Les campagnes menées ont révélé des preuves concrètes de pratiques problématiques : discriminations tarifaires sur AliExpress (variations de prix jusqu'à 183%), orientations biaisées des résultats de recherche sur Amazon (écarts moyens de prix allant de 235€ à 1100€), et promotions trompeuses sur Temu. Ces résultats soulignent l'importance d'une surveillance continue des plateformes numériques.

Les travaux futurs viseront à étendre la plateforme aux écosystèmes iOS et Web, à améliorer l'extraction automatique de données via des modèles de vision LLM, et à faciliter l'analyse des résultats grâce à des outils de visualisation comme Power BI. L'objectif à long terme est de rendre STETOSCOPE accessible au grand public pour une meilleure transparence algorithmique.

Références

Références

- [1] Beaufls, T., Boutet, A., et al. *STETOSCOPE : underStand TargEting and manipulaTiOnS via Collabo-rative Private data collection*. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.
- [2] Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). *The dark (patterns) side of UX design*. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
- [3] Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). *Dark patterns at scale : Findings from a crawl of 11K shopping websites*. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-32.
- [4] Hannak, A., Soeller, G., Lazer, D., Mislove, A., & Wilson, C. (2014). *Measuring price discrimination and steering on e-commerce web sites*. In Proceedings of the 2014 conference on internet measurement conference (pp. 305-318).
- [5] Regulation (EU) 2016/679 of the European Parliament and of the Council. *General Data Protection Regulation (GDPR)*. Official Journal of the European Union, 2016.