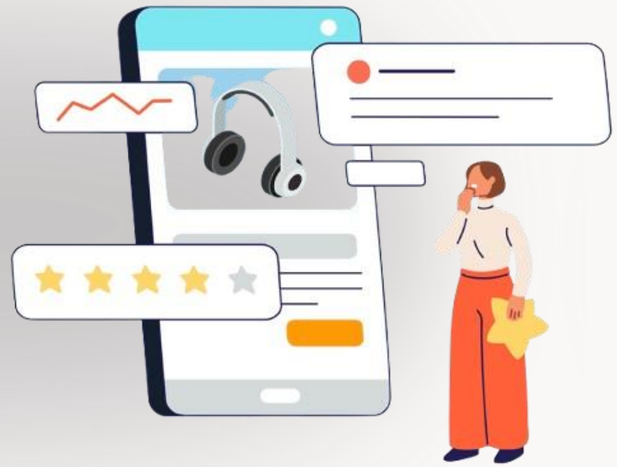
AI and Fake Reviews
amazon's Struggle Against Bots



Presented by:
Group 10

Submitted to:
Dr. Susan Leavy

Group Members

Student ID	Student Name
24298473	Yuan yuan Xiang
24223960	Parijat Dhar
24295848	Sai Sugun D R
24202473	Nan Zhang
24234516	Shivani Singh
24230224	Bhuvana Kodathi Srinivas
24238836	Dhanush Devadiga
24216949	Yathish Kumar Surendra Kumar
24216940	Vaishnavi Kandikonda

A bit of Context.

“Guerrilla marketing strategy”

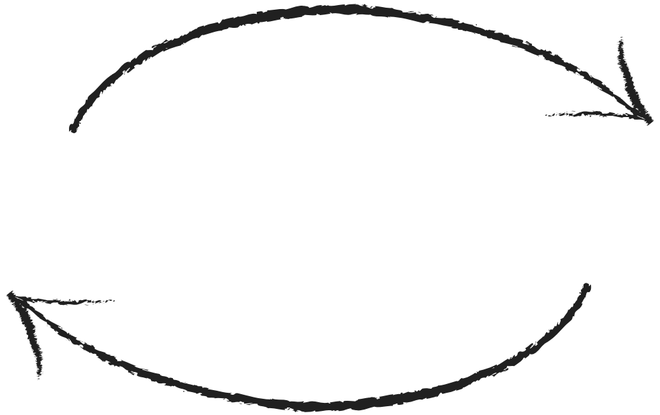
reddit

“Seeding fake profiles”

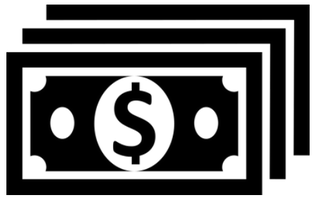
“What” & “How”

Reviews / Ratings

AI Bots

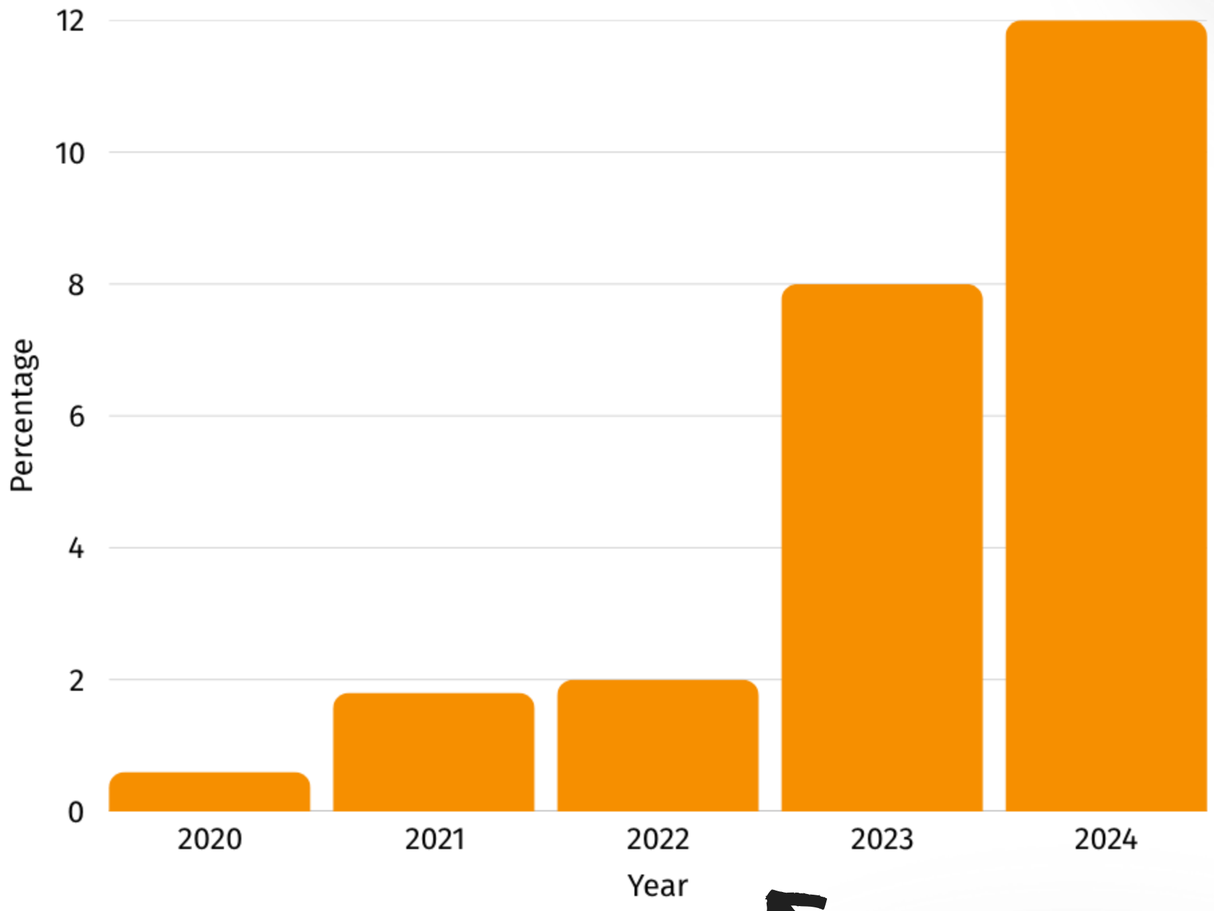


Increased sales of Low Quality Products



Incentives

AI Generated Reviews on Amazon



OpenAI ChatGPT 4.0

(Originality.ai, 2024)

Effects & Impacts

Unfair Competitive Advantage

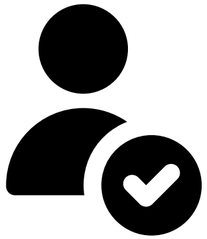
Exploitation of Trust

Negative Brand Impact

Accountability & Transparency Issues

AI engineers' perspective

Amazon’s Existing System – Overview

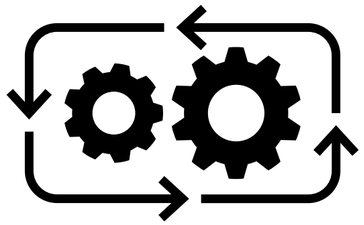


Verified Purchase Tags

1

2

Machine Learning-Based Detection



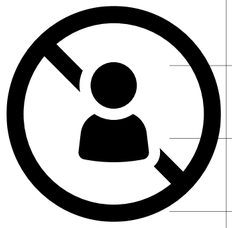
3

Behavioral & Historical Analysis

4

Manual & Crowd sourced Moderation

5

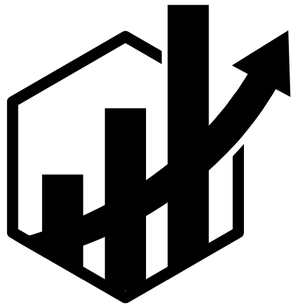


Suspicious Review Removal & Bans

Amazon’s Existing System – Limitations

- INEFFECTIVE AGAINST AI-GENERATED REVIEWS
- NO REAL-TIME NETWORK GRAPH ANALYSIS
- LACK OF TRANSPARENCY & EXPLAINABILITY
- NO LABELING FOR AI-GENERATED CONTENT
- REACTIVE INSTEAD OF PROACTIVE DETECTION
- NO MULTIMODAL (IMAGE/VIDEO) REVIEW VALIDATION

Adaptive Multi-Layer System for Fake Review Detection



Behavioral Trust Graph (BTG)

Reviewer Identity Fingerprinting

Ethical Labeling & Trust Tags

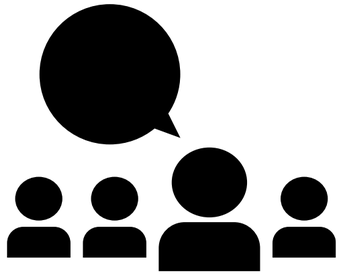
1 — 2 — 3 — 4 — 5 — 6

1

LLM-Adaptive Detection

3

Sentiment & Intent Validation



5

Explainable AI for Moderators

Consumer Protection Groups' perspective

Scale of the Problem

200M+

fake reviews blocked by Amazon in 2022 alone

(Bright local's, 2023)

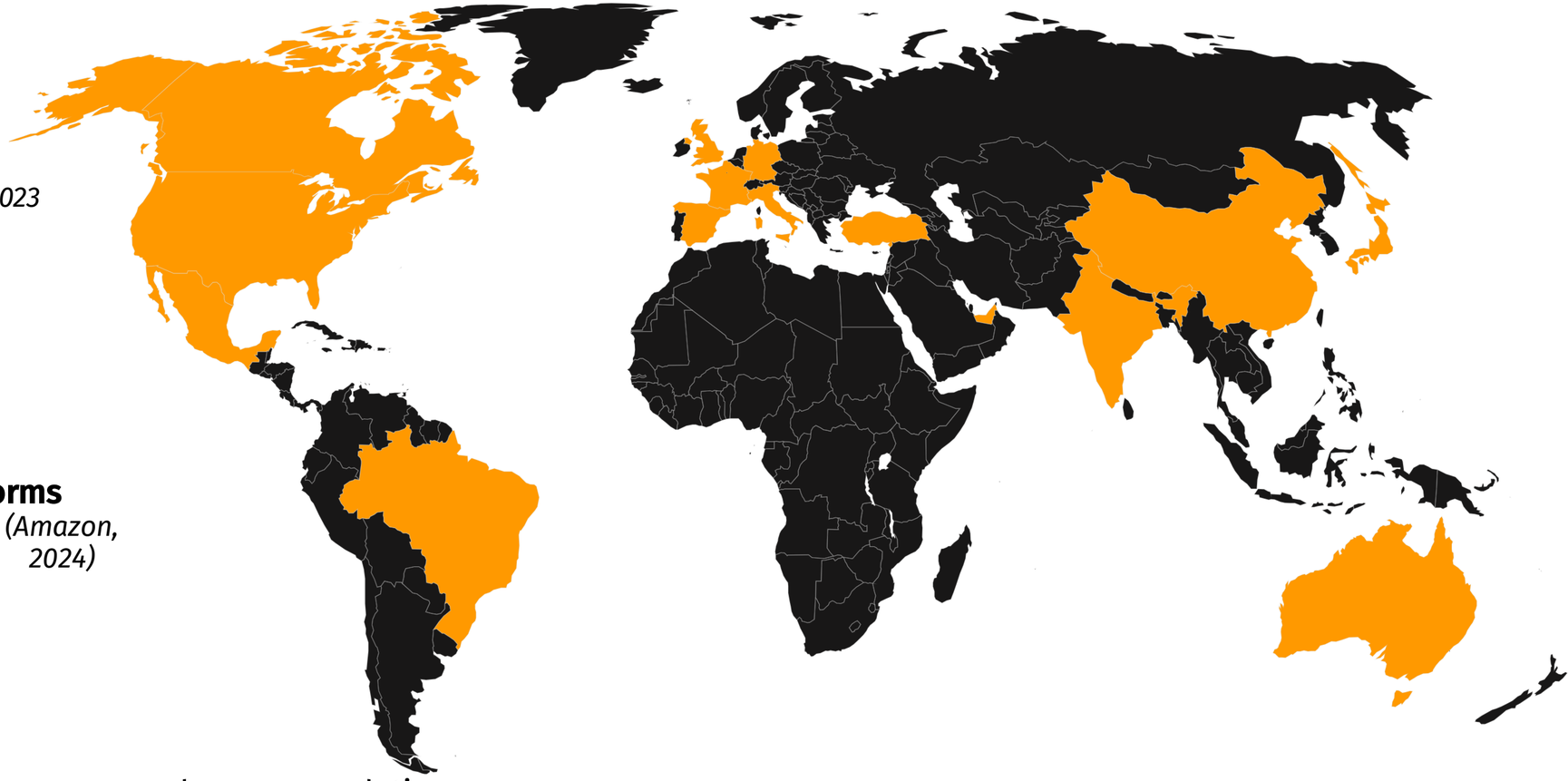
91%

fake reviews blocked by Amazon in 2022 alone

20 - 40%

reviews are likely to be fake on many online platforms

(Amazon, 2024)



Key Concerns:

- ***Trust Erosion:*** 76% trust online reviews as much as personal recommendations.
- ***Economic Harm:*** Billions in annual consumer damage (FTC estimates)
- ***Information Asymmetry:*** Consumers can't distinguish AL- generated from authentic reviews.
- ***Vulnerable Populations:*** Elderly and digitally inexperienced consumers are most affected.

Consumer Protection Strategy

Four-Pillar Approach

Transparency

- AI content labelling for pre-existing reviews
- Enhanced Verification Systems
- Public Authencitity Scoring

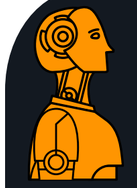
Education

- Awareness Campaigns
- Enhanced Verification Systems
- Browser tools for pattern flagging

Accountability

- Regular Platform Audits
- Legal Consequences for Sellers
- Enhanced class-action rights

Technology



- AI powered detection systems
- Statistical Analysis
- Advanced language processing

Success Stories & Ongoing Challenges

Leading Organizations

- Consumer Reports: Educational campaigns and platform accountability.
- Consumer Federaton of America: FTC testimony and refulatory enforcement.
- UK Which? : Successfully influenced regulatory approaches to fake reviews
- CHOICE (Australia): Consumer guidance and accountability framework.

Current Challenges

- *Technological Arms Race: AI vs detection capabilities.*
- *Balancing Act: Authenticity without penalizing legitimate business.*
- *Accessibility: Ensuring verfication doesn't create barriers.*
- *Scale: Platforms removing millions of fake reviews annually*

16 %

increase in fake review detection (Trustpilot 2021 -2022)

5 -9 %

revenue increase per 1-star rating boost

Government Officials' perspective

Legislative perspective: Existing Legal Frameworks

FTC Act Section 5	PROHIBITS UNFAIR OR DECEPTIVE ACTS OR PRACTICES IN COMMERCE.
EU Digital Services Act	REQUIRES PLATFORMS TO ADDRESS RISKS OF MANIPULATED CONTENT.
Consumer protection laws	IN VARIOUS JURISDICTIONS THAT ADDRESS FALSE ADVERTISING.
Limited specific legislation	TARGETING FAKE REVIEWS IN MOST COUNTRIES.
Critical gap	MOST LAWS WERE WRITTEN BEFORE AI-GENERATED CONTENT BECAME SOPHISTICATED.

Supervisory Perspective: Direct Regulation of AI-Generated Reviews

Prohibitions and Penalties

- Criminal penalties for commercial fake review services.
- Fines are scaled to company size for beneficiaries of fake reviews.
- Bans on AI tools specifically marketed for review manipulation.
- Rewards/protections for whistleblowers reporting violations.

Technical Standards

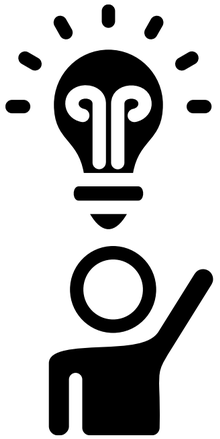
- Developing certification standards for review authentication systems, which is ISO 20488:2018.
- Requirements for AI watermarking or fingerprinting in content generation tools.
- Technical guidelines for minimum review verification standards.
- International standards for cross-border enforcement cooperation.

Proposed Regulatory Approaches

Collaborations with Platforms

1. Disclose review screening criteria.
2. Leverage AI for fake review detection.
3. Purchase verification system.

Enforcement Challenges



1. Cross-border jurisdiction issues arise when review firms operate internationally.
2. Difficult to attribute AI-generated content to specific actors.
3. Limited technical expertise within regulatory agencies.
4. Resource constraints for monitoring vast review ecosystems.
5. Evidentiary standards that may not accommodate AI detection methods.

Thanks!

Any Questions?

Sources and References

- Amazon. (2023). *2022 Brand Protection Report*. Retrieved from amazon.com/brandprotection
- BrightLocal. (2023). *Local Consumer Review Survey 2023*. Retrieved from brightlocal.com/research
- CHOICE. (2023). *Online Reviews and Consumer Trust*. Retrieved from choice.com.au
- Consumer Federation of America. (2022). *Testimony on Fake Reviews to FTC*. Retrieved from consumerfed.org
- Consumer Reports. (2023). *Fake Reviews Investigation Reports*. Retrieved from consumerreports.org
- Federal Trade Commission. (2022). *Fake Reviews and Other Misleading Endorsements*. Retrieved from ftc.gov
- He, S., Hollenbeck, B., & Proserpio, D. (2022). The Market for Fake Reviews. *Marketing Science*, 41(5), 896-921.
- Luca, M. (2016). Reviews, Reputation, and Revenue: The Case of Yelp.com. *American Economic Journal: Applied Economics*, 8(3), 136-153.
- Trustpilot. (2023). *Transparency Report 2022*. Retrieved from trustpilot.com/trust/transparency-report
- Which?. (2023). *Fake Reviews Campaign Reports*. Retrieved from which.co.uk
- Fornaciari, T., & Poesio, M. (2014). Identifying fake Amazon reviews by learning from crowds. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (pp. 279-287). Association for Computational Linguistics.
- Amazon Newsroom (2023) – How Amazon Uses AI to Spot Fake Reviews https://www.aboutamazon.com/news/policy-news-views/how-ai-spots-fake-reviews-amazon.
- Allows civil penalties up to $50,000 per fake review. Federal Trade Commission (FTC) press releases and proposals.
- Classifies deceptive AI tools as "high-risk"; subject to compliance and bans. European Commission - EU AI Act final draft 2024
- Machine learning detects and blocks ~25% of fake reviews. Amazon Transparency Report / Yelp Trust & Safety updates