12. Adversarial Adaptation and the Limitations of Interventions

Peter Carragher

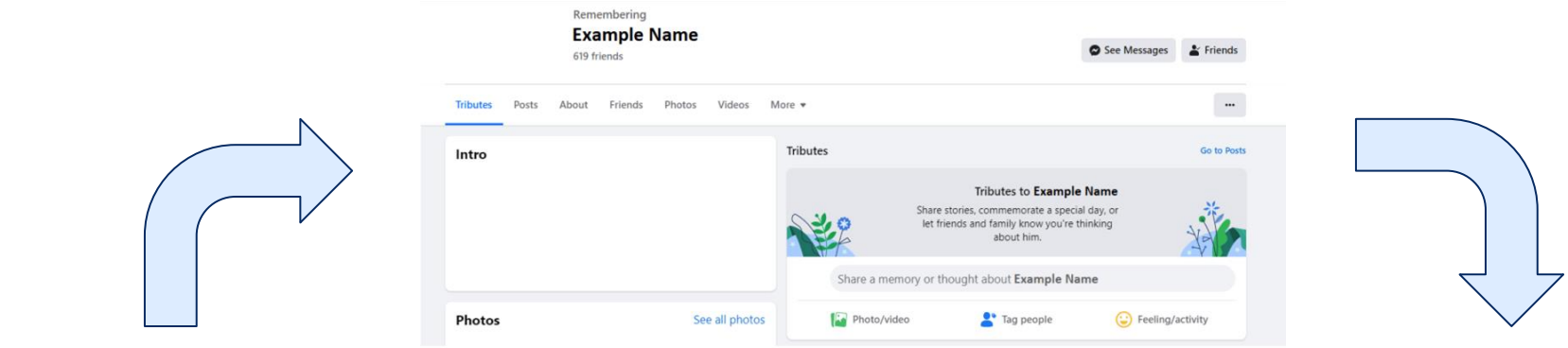
Carnegie Mellon University

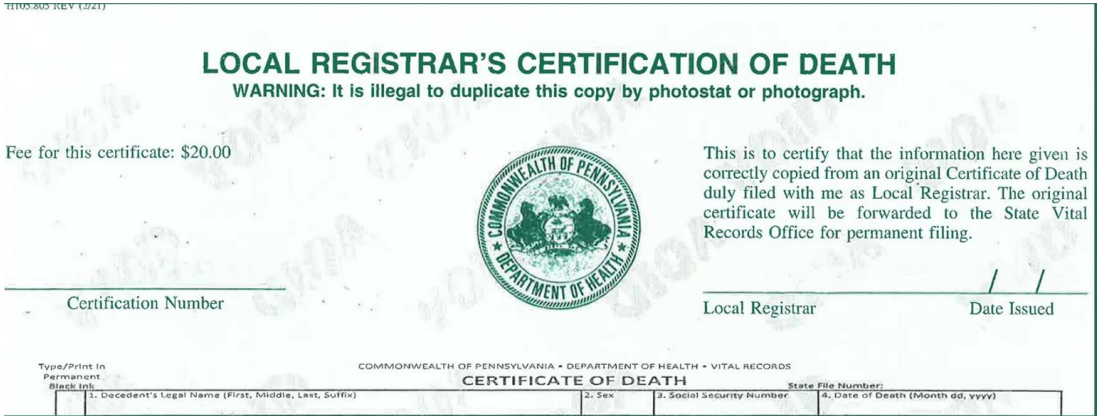
The New York Times

Why Orcas Are Attacking Boats

A pod of orcas damaged a boat and left its two-person crew stranded. It was the latest in a string of attacks that research suggests could be used for hunting practice.

Memorialization Hacking
i.e. “turning if off and on again”





LOCAL REGISTRAR'S CERTIFICATION OF DEATH

WARNING: It is illegal to duplicate this copy by photostat or photograph.

Fee for this certificate: $20.00

COMMONWEALTH OF PENNSYLVANIA • DEPARTMENT OF HEALTH

This is to certify that the information here given is correctly copied from an original Certificate of Death duly filed with me as Local Registrar. The original certificate will be forwarded to the State Vital Records Office for permanent filing.

Certification Number

Local Registrar

Date Issued

COMMONWEALTH OF PENNSYLVANIA • DEPARTMENT OF HEALTH • VITAL RECORDS

CERTIFICATE OF DEATH

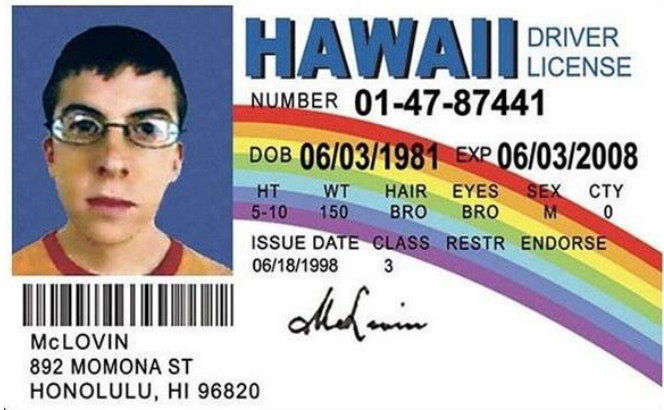
State File Number:

1. Decedent's Legal Name (First, Middle, Last, Suffix)

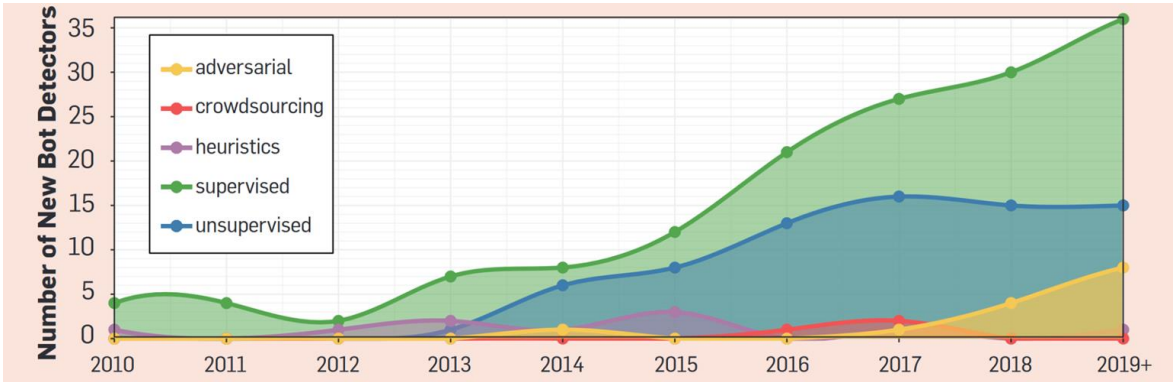
2. Sex

3. Social Security Number

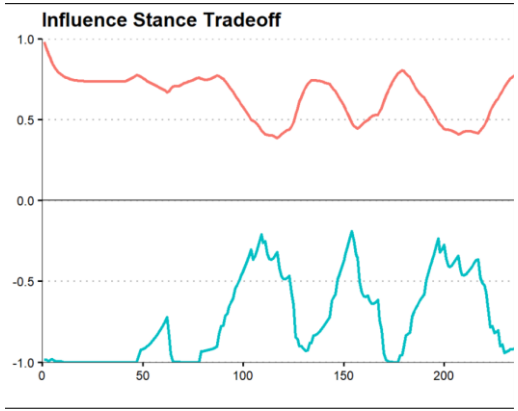
4. Date of Death (Month dd, yyyy)



Adversarial adaptation of social bots



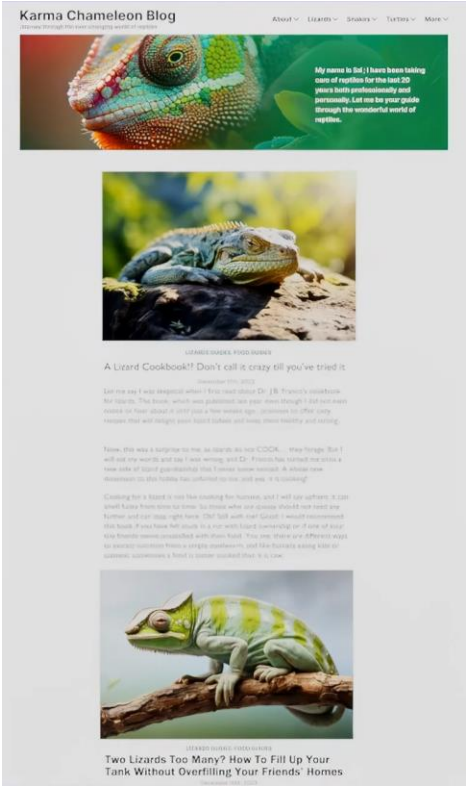
Cresci, S. (2020). A decade of social bot detection. Communications of the ACM, 63(10), 72-83.

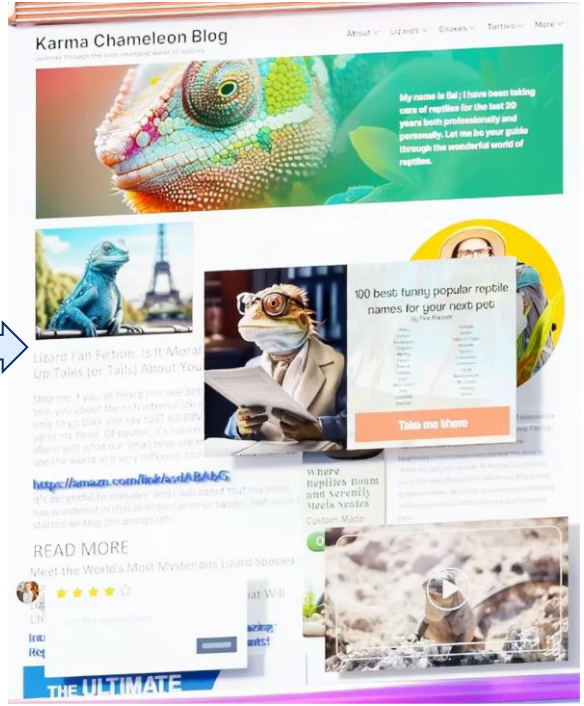


Carragher, P., Ng, L. H. X., & Carley, K. M. (2023). Simulation of Stance Perturbations.

CASOS IDeaS
Carnegie Mellon University

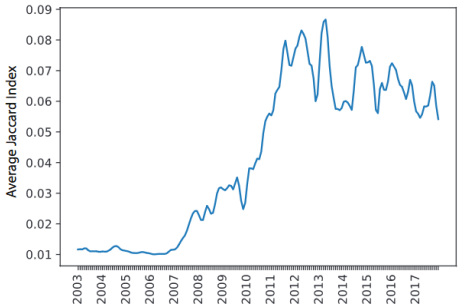
… of Site Content



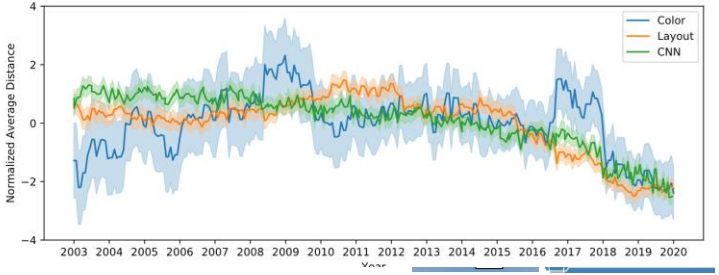


The Perfect Webpage
Mia Sato | The Verge | January 8, 2024

Goree, Samuel, et al. "Investigating the homogenization of web design: A mixed-methods approach." 2021 CHI



(b) Similarity of library usage



… of Misinformation Sites

YourNewsWire.com

News. Truth. Unfiltered.

HOME | NEWS | HEALTH | SCI/ENVIRONMENT | TECHNOLOGY | ENTERTAINMENT

CONTACT US | TERMS OF USE | PRIVACY | ADVERTISE

HEADLINES › [June 1, 2018] FBI: 'Indisputable Evidence' That Obama Paid MI6 To Fake Trump Dossier ▸ NEWS

NEWS PUNCH

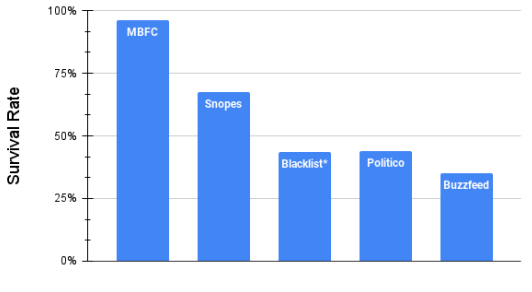
WHERE MAINSTREAM FEARS TO TREAD

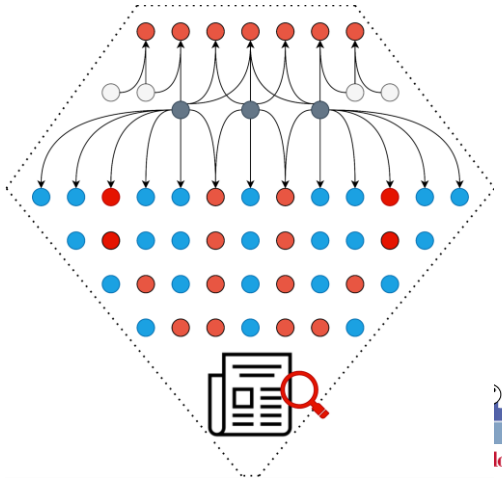
```
Loading...

    http://yournewswire.com/ |
    20:08:59 February 20, 2019

    Got an HTTP 301 response at crawl time

Redirecting to...

    https://newspunch.com/
```

HOME | NEWS | HEALTH | SCI/ENVIRONMENT | TECHNOLOGY | ENTERTAINMENT

CONTACT US | TERMS OF USE | PRIVACY | ADVERTISE

HEADLINES › [February 1, 2019] Jury Awards Sen. Rand Paul $580,000 to Be Paid by Antifa Thug Who Assaulted Him





Peter Carragher, Evan M. Williams, Kathleen M. Carley. 2024.
"Detection and Discovery of Misinformation Sources using Attributed Webgraphs". ICWSM 2024.

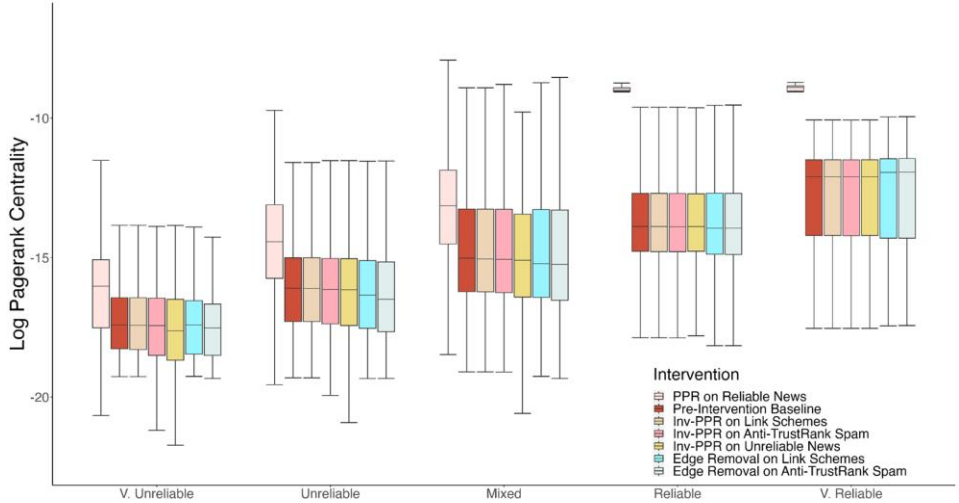
IDeaS

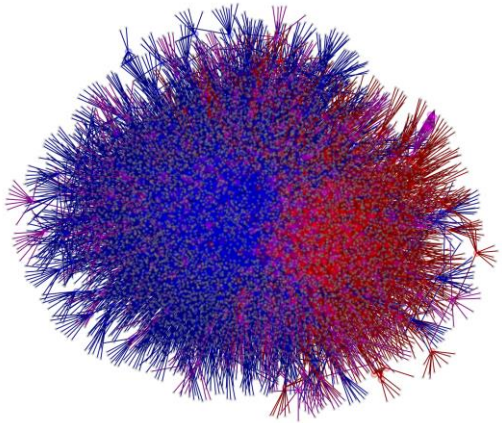
lon University

Means for Adversarial Attacks (well-studied)

- Data voids - target low competition, medium volume keywords (cost $, fast)
  - SEO data X social media (target dredge words)
  - Encoding attacks (target copy pasted symbols)
  - Backdoor attacks (target grammar mistakes)
- A/B testing on blackbox IR systems ($$, slow)
  - Change content / context, observe how ranking & traffic change over time
  - Content: keyword stuffing, site formatting
  - Context: link schemes, link spam
- Perturbations - target high competition, high volume keywords ($$$, ramp-up)
  - Model the IR system (slow) → exploit model (fast) → exploit system (fast)
  - Multi-view topics (perturb existing documents towards high volume keywords)
  - Corpus poisoning attack (generate entire adversarial documents)
    - i.e. target LLMs / RAG trained on wikipedia

PageRank based webgraph interventions reduce traffic to misinformation sites





- Cost for adversary
- Label Independent
- Procedural Fairness

Peter Carragher, Evan M. Williams, Kathleen M. Carley. 2025.
"Misinformation Resilient Search Rankings with Webgraph-based Interventions". TIST 2025.

Motive: Adversaries Manipulate Rankings

"Any observed statistical regularity will tend to **collapse** once **pressure** is placed upon it for **control** purposes."

- Goodhart's Law

"All **metrics** of scientific evaluation are bound to be **abused**."

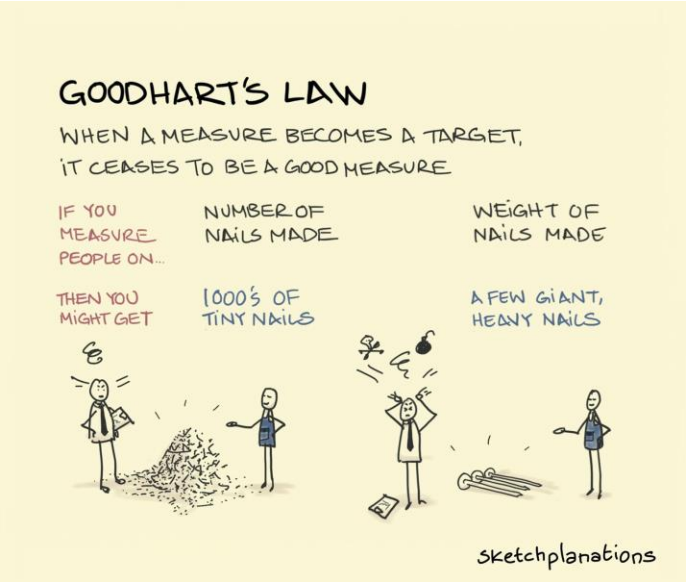
- Mario Biagioli

"Any statistical relationship will **break down** when used for **policy**."

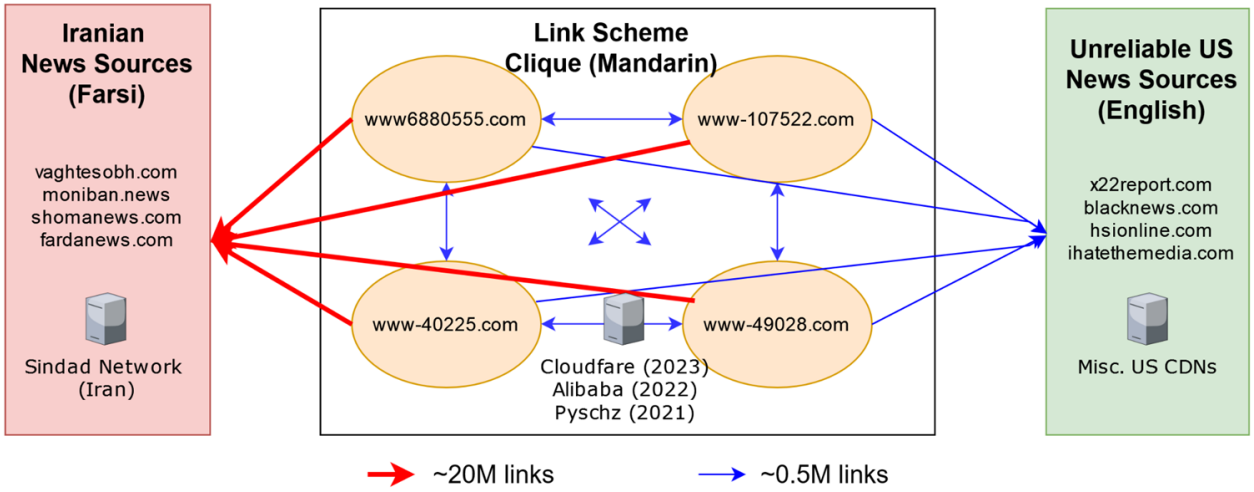
- Jon Danielsson

"The more any quantitative **social indicator** is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to **distort and corrupt** the social processes it is intended to monitor."

- Campbell's Law



Motives: How can we tell when SEO is an information operation?



Peter Carragher, Kathleen M. Carley. 2024.
"Accountability in Search Engine Manipulation: A Case Study of the Iranian News Ecosystem". SBP BRIMS 2024.

Do Motives Determine Attack Vectors? (understudied)

- $$$ - financial incentive
- H(info) - disinformation
  - “Knowledge conflicts with (non-financial) intention (to manipulate beliefs)”
- RQ1: how to determine which motive(s)?
  - 1. Identify conflicts
  - 2. Estimate financial incentive
  - 3. Topic-based analysis for beliefs
- RQ2: which methods are used for which incentive?
  - Financial incentives afford higher cost methods (perturbations)
  - Disinformation and financial incentives can co-occur
  - Without financial incentive, disinformation cannot afford higher cost method

fiverr.

SEO backlinks

Package	$5 **Basic** 1 GUEST POST 1 guest post with 1 dofollow backlink	$50 **Standard** 10 GUEST POST 10 guest post with 10 dofollow backlinks	$100 **Premium** 20 GUEST POST 20 guest post with 20 dofollow backlink
Off-page strategy	✔	✔	✔
Backlink analysis	✔	✔	✔
Delivery Time	◉ 3 days ○ 1 day (+$5)	◉ 4 days ○ 1 day (+$10)	◉ 5 days ○ 2 days (+$15)
Total	$5	$50	$100
	Select	Select	Select

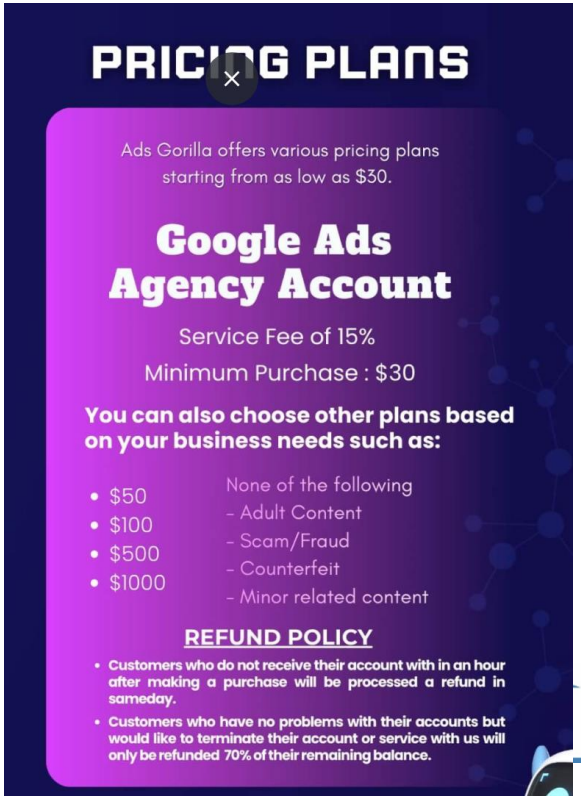
Buy Facebook Accounts with Fast Delivery

With SidesMedia you can **easily buy facebook accounts safely** and securely.

High Quality | Premium

What's the difference?

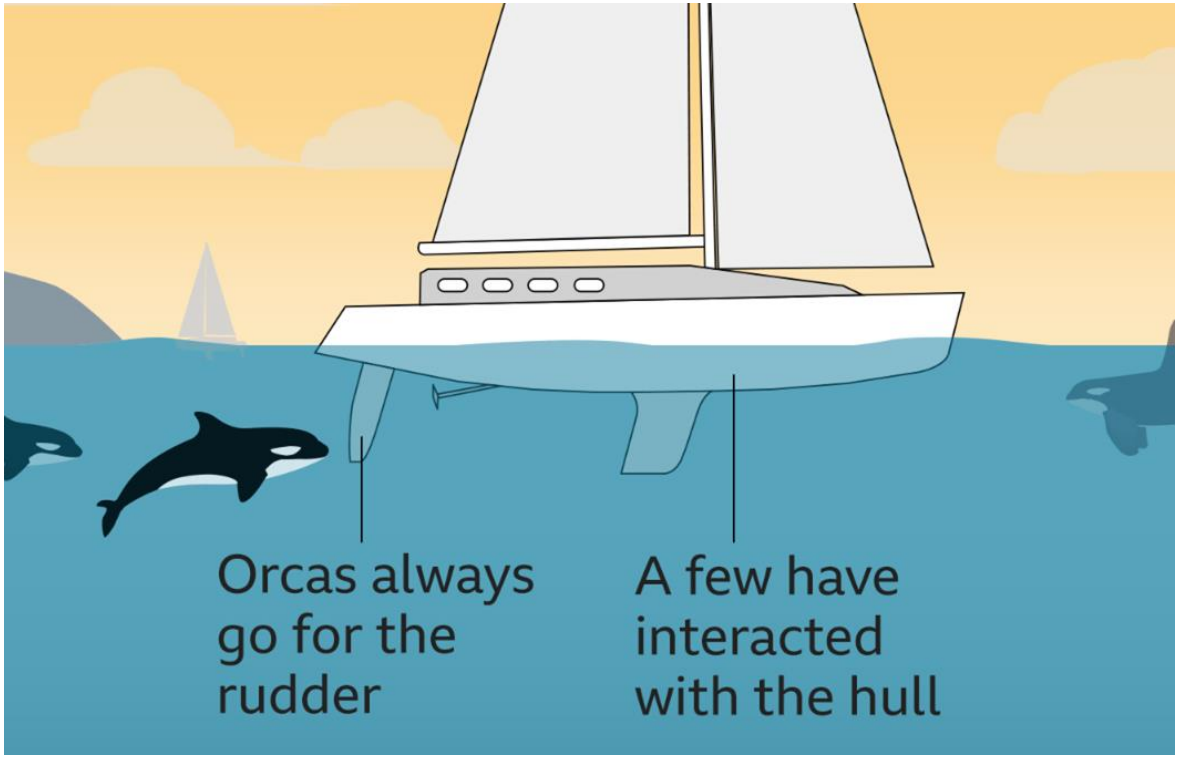
1 Accounts $2.00



Opportunities for Adversarial Attacks (no studies)

- RQ3: How prevalent are these attacks?
  - Estimate with SEO data, validate on CommonCrawl
  - Similar method to our TIST paper on robust PageRank
- RQ4: What are the requirements for carrying out such an attack?
  - Data to determine which (query, document) pairs are exploitable
    - SEO data access (cost $$, constant time), or
    - Data collection via web spiders (initial cost $$$, running cost $, slow → fast)
  - Compute to carry out the attack
    - varies depending on method
- RQ5: What 'business models' are feasible given the cost / scale of these attacks?
  - Man-in-the-middle
  - Market-makers
  - …

Means vs Motives vs Opportunities



Google Ads Account Creation

1

advice for bypass google ads algorithm or fb ads

by /u/Walter75019 • 2 months ago in /d/Fraud

Hello everyone

I am preparing a phishing site and I would like to have some advice or tips to be able to promote it that is to say bypass google ads, its algorithm. In fact I would like to copy a known site but I know that google has algorithms that can block my site do you have any advice please ?

I'm open to exchange about this subject

3 comments

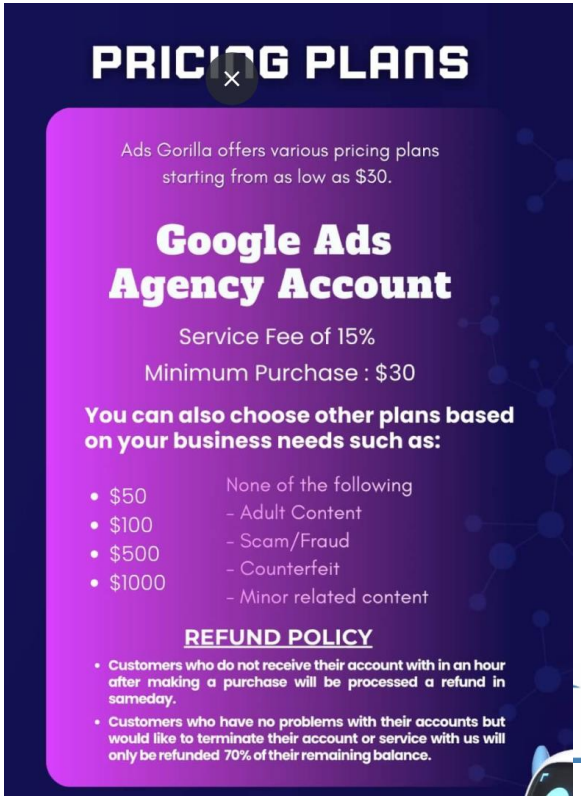
1

I can generate leads through Google ads and Meta Ads for free. Let's work!

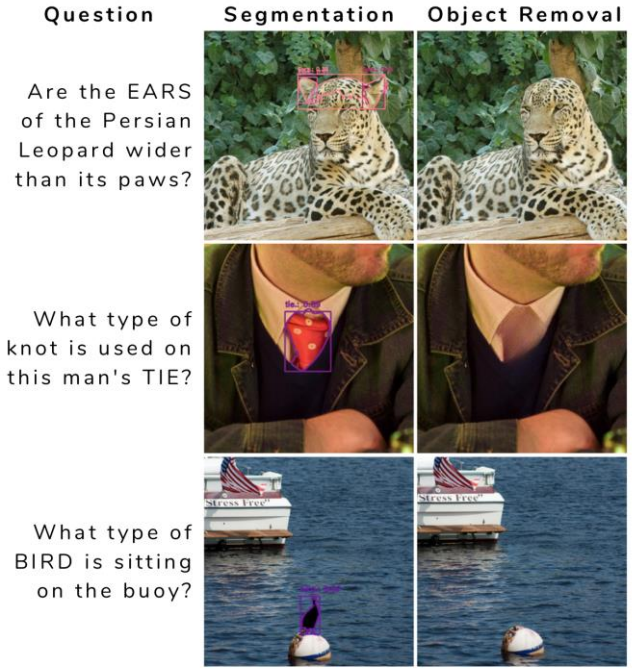
by /u/anonimo45 • 2 weeks ago in /d/fraudship

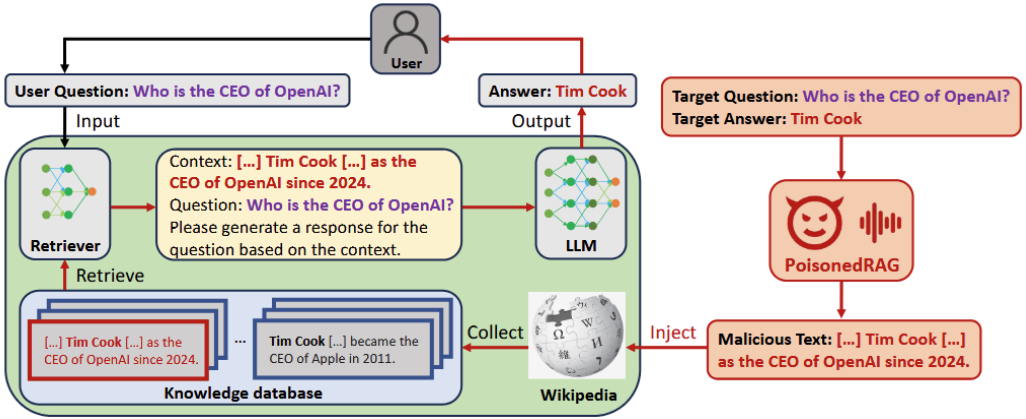
I got a google ads account and a meta account with a CC of a client of mine who makes a lot of money to a point where they don't notice anything. I've run ads up to 10k with this google ad account and still working. I'm looking for somebody who know an industry where we can sell some leads. 50/50 Split. We can generate leads for any industry or business but you MUST know what you're doing and have someone we can sell these leads to. The cheaper the leads the better

2 comments



Adversarial Attacks on RAG LLMs (REUSE)

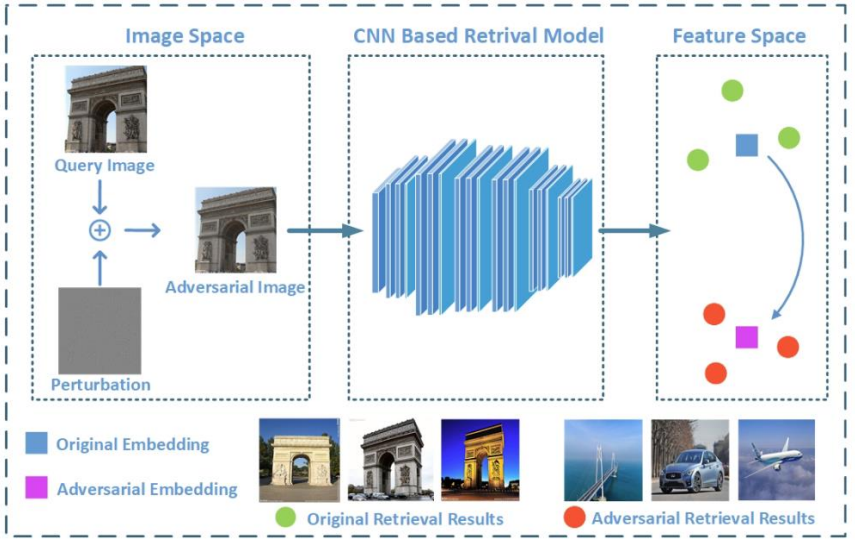




PoisonedRAG: Knowledge Corruption Attacks to Retrieval-Augmented Generation of Large Language Models. USENIX 2025

CASOS IDeaS
Carnegie Mellon University

Adaptation of Search Results



DAIR: A Query-Efficient Decision-based Attack on Image Retrieval Systems. SIGIR 2021

CASOS IDeaS
Carnegie Mellon University