FENIMORE HARPER | Say No To Disinfo

Can A.I. Cause a Bank Run?

- **Say No to Disinfo** and **Fenimore Harper Communications** **simulated an AI-powered disinformation campaign**, with the goal of causing a bank run in the United Kingdom.

- When exposed to the campaign, **60.8% of polled individuals expressed a likelihood of moving their money**, revealing the potent ability of synthetic content to incite financial instability.

- For every £10 spent on ads, **as much as £1 million could be moved**. This demonstrates the extraordinary cost-effectiveness of AI-driven influence operations in triggering bank runs.

- Banks which focus on cyber threats and **neglect the risks posed by influence operations**, including a **lack of disinformation expertise, threat mapping, and war gaming, leaving them critically vulnerable**.

‘Can AI Cause a Bank Run?’
CONTENTS

Executive Summary

This report looks at the potential for AI augmented influence operations to trigger bank runs through targeted disinformation campaigns. As AI is making disinformation campaigns easier, cheaper, quicker and more effective than ever before, the emerging risk to the financial sector is rapidly growing but often overlooked. As the cost and speed of creating and spreading disinformation campaigns is plummeting, it is no longer solely the domain of large state or non-state actors, there may be a wider range of actors with motives ranging from ideology and financial gain to political impact that could look to target the sector. With the rise of online banking and widespread nature of secondary accounts, it is becoming increasingly quick and easy for customers to move money between their accounts. With asymmetric cost functions when there may be a perceived risk of losing money, banks are increasingly vulnerable.

To explore the potential impact of such influence operations on the financial sector, we used A.I. to generate a ‘fake news’ campaign targeting the financial health of banks, coupled with polling on customer behaviour change triggered by the campaign.

Using A.I. tools we generated false headlines focused on narratives tapping into existing fears or biases, with a key message of: “customer money is not safe”. To mimic the spread of these operations using doppelganger websites, we simulated similar sites which look like trusted sources. Using X/Twitter amplification to add ‘social proof’ and mimic social media spread, we generated posts and humour based memes at scale, finding that 1,000 tweets can be generated in less than a minute, at a trivial cost.

To evaluate the impact of a campaign, we polled a random cohort of 500 people across the UK, showing them the content generated. Our findings show that after having been shown the synthetic content 33.6% of people are extremely likely and 27.2% of people somewhat likely to move their money, and 60% of people would share this with 1-3 people,

20% with more than 3 people. Assuming 80% and 50% respectively of these groups would move their money, we estimate 405 customers moving money for 1,000 ads shown.

Using the average UK bank account, we estimate the amount of money moved would be £3 million before considering sharing behaviour, and £10 million if we incorporate sharing numbers.

According to Revolut's annual report, they hold c.£5bn in capital, and have made £15.1bn in customer deposits. Using the average cost of ads across social media platforms including X, Facebook, Youtube and Tiktok, we estimate that **to move 1% of total loans (£150M) would cost $90-$150, to move 5% would cost $450-$750 and to move 30% would cost $2,700-$4,500.**

Given the speed, ease and trivial cost at which an effective influence operation can be run, the financial sector needs to be prepared. However financial institutions lack disinformation specialists, rarely have done any trust mapping for customers, rogue actor mapping or war gaming and tend to often be reactive rather than proactive.

Current assessments conducted by banks tend to focus on operations, systems and cyber threats and neglect the ways in which AI-enhanced influence operations could affect their customers. It is critical that banks understand this risk, and their underlying vulnerabilities that could be exploited. Our report sets out the key recommendations for banks and regulators to consider to build a more resilient financial sector.

Introduction

In an AI age, disinformation is increasingly quicker and cheaper to create and spread. As content creation is becoming more automated, this reduces the financial and time costs associated with micro targeting and hyper personalization, and an improved understanding of the information environment allows threat actors to craft more compelling and effective narratives for each target segment. The use of increasingly realistic deepfake profile photos, with LLM generated bio's and online presence, can enable en masse creation of credible accounts to spread disinformation.

The speed of social media and the ease of online banking mean that false narratives around the solvency of a bank or other relevant claims, can go viral and trigger a large-scale customer response and even a bank run. This could also be spread to multiple banks and there is a risk of financial contagion. These emerging risks posed by disinformation to financial institutions and the financial sector as a whole are not well understood or quantified.

This report looks at the potential for targeted AI-augmented information campaigns designed to instigate a series of bank runs, causing widespread financial instability and economic damage. Using polling across customer behaviours, the use of AI tools for content creation, and cost estimates for ads across different social media platforms, we simulate false content creation and estimate the cost to trigger a bank run using AI-enhanced disinformation.

Part 1: Context: Bank Runs

1. What causes a bank run?

A bank run is caused when a large number of the bank's customers withdraw funds (or transfer money to a different financial institution) from the bank, due to the belief that the bank may fail and their money is not safe. As more and more withdrawals occur, this causes the likelihood of default to increase, which further encourages more customers to make withdrawals, causing a self fulfilling prophecy. This can destabilise a bank to the point of bankruptcy.

2. What are some of the recent bank runs / liquidity crises?

Bank runs from 2008 and more recently SVB/First Republic Bank keep the risk of bank runs in the public's psyche, raising the profile of the perceived risk. The three bank failures in 2023 fall far short of the several dozen failing every year between 2008 and 2013, however 2023 is by far the most costly year ever for bank failures measured by the shuttered institution's assets, according to LPL Financial research.

2019 - Metro Bank

In 2019 a false story was spread via Whatsapp within the Tamil community in West London, claiming that Metro Bank was facing financial difficulties and may be shut down. As this continued to spread rapidly, queues started to form outside branches, of customers fearful of a collapse wanting to move money out of their accounts, or empty their safe deposit boxes. Many photos on Twitter show crowds of people waiting in line at various branches.

Metro Bank representatives quickly came out with statements refuting this, saying "We're aware there were increased queries in some stores about safe deposit boxes following false rumours about Metro Bank on social media and messaging apps, there is no truth to these rumours and we want to reassure our customers that there is no reason to be concerned." However recent news stories, which may have been twisted

on social media, about the financial regulator raising concerns about accountancy errors in bank reports, and that Metro Bank had miscalculated how much capital it needed to back up its commercial lending operations, fuelled the narrative. Despite managing to continue operations, the bank has lost up to 24% of its retail customers.

2023 - Silicon Valley Bank - the first social media fuelled bank run?

SVB was the second largest bank failure in the United States. The largest was Washington Mutual Bank in 2008, when $16.7 billion was withdrawn over the course of 10 days, in comparison a reported $4.2 billion was withdrawn from SVB in just 24 hours in March 2023. Congressman Patrick McHenry, chairman of the US House Financial Services Committee, referred to the turmoil as, “the first Twitter fuelled bank run.” The speed of social media and the ease of online banking mean that narratives around the solvency of a bank can go viral and trigger a large-scale customer response. After SVB’s announcement of a decision to raise funds through a sale of shares, triggering online commentary such as Silicon Valley investor Jason Calacanis tweeting “YOU SHOULD BE ABSOLUTELY TERRIFIED RIGHT NOW” in a tweet that has had more than 4.9 million views. In the absence of a counternarrative from the bank, customers relied on online information, spreading panic throughout social platforms on talk of a bank run, resulting in a self-fulfilling prophecy.

2023 - First Republic Bank

Data from Valent Projects reveals that First Republic Bank, the second largest American bank to fail in history, was targeted by an online manipulation campaign. An extensive network of bots and fake accounts was identified, which was actively steering the social media discourse on the topic, aggressively amplifying negative narratives. This coincided with a large increase in short positions against the bank. Just a few months later, the bank collapsed.

The bank knew that social media had been filled with concern over their future after Silicon Valley Bank collapsed; First Republic Bank,

another regional bank in the United States, immediately began to feel the ripple effect, creating a wave of panic. The withdrawal of $40bn from SVB in one day in March had already demonstrated the dizzying speed at which deposits can be taken out in the digital banking age, and how social media can amplify a panic. First Republic was more diversified than SVB's heavily tech sector-dominated client base. The bank was facing a couple of years of poor earnings, but might well have survived had it not suffered a run on its deposits.

3. Who might want to cause a bank run and why?

There is an increasing range of potential threat actors who may want to cause a bank run with different motives.

Financial Gain:

Financial gain may be a significant motivating factor for actors who could run influence operations. Commercial or financial actors that may be shorting the bank could target them to make a financial gain. Potentially even competitor banks could target a bank through proxies that make it easier to avoid detection. Dark PR firms could be used by these actors to design and implement campaigns.

Ideological Motivation:

Disgruntled figures such as ex-employees could launch a campaign with a modest set of resources, using AI to automate large parts of the influence campaign. Activist groups that disagree with bank policy (e.g. on climate change or net zero) could use campaigns as a tool to effect change by disrupting the bank's operations.

Political Impact:

There is a risk of terrorist groups and hostile foreign states who may target banks to create chaos, economic damage and even political stability. Banks are a critical element of critical infrastructure, which makes them a critical target. As was seen in 2008, the impacts

of a “credit crunch” can be hugely damaging and long-lasting. At the most simple level, without access to credit, people and businesses are unable to function for long. As was seen in 2008, a run on a retail bank can have a hugely damaging impact on individuals but it also threatens macroeconomic stability, destabilise the economy and causes long term damage to a country.

Part 2: Context: Misinformation Campaigns / Influence Operations

1. How do influence operations work?

Content creation, distribution and amplification designed to shape views/opinions. This could be creating synthetic content, fake accounts or even just amplifying polarising content or that supports a particular viewpoint. These are observed across a wide range of topics ranging from trust in science/government through to consumer finance.

a. How many channels do they use?

Both offline and online. Offline examples may include print media, television, public billboards etc. Online could be via ads, e.g. search, social media content, or even in forums or chatbots as well as doppelganger websites.

b. What strategies are implemented?

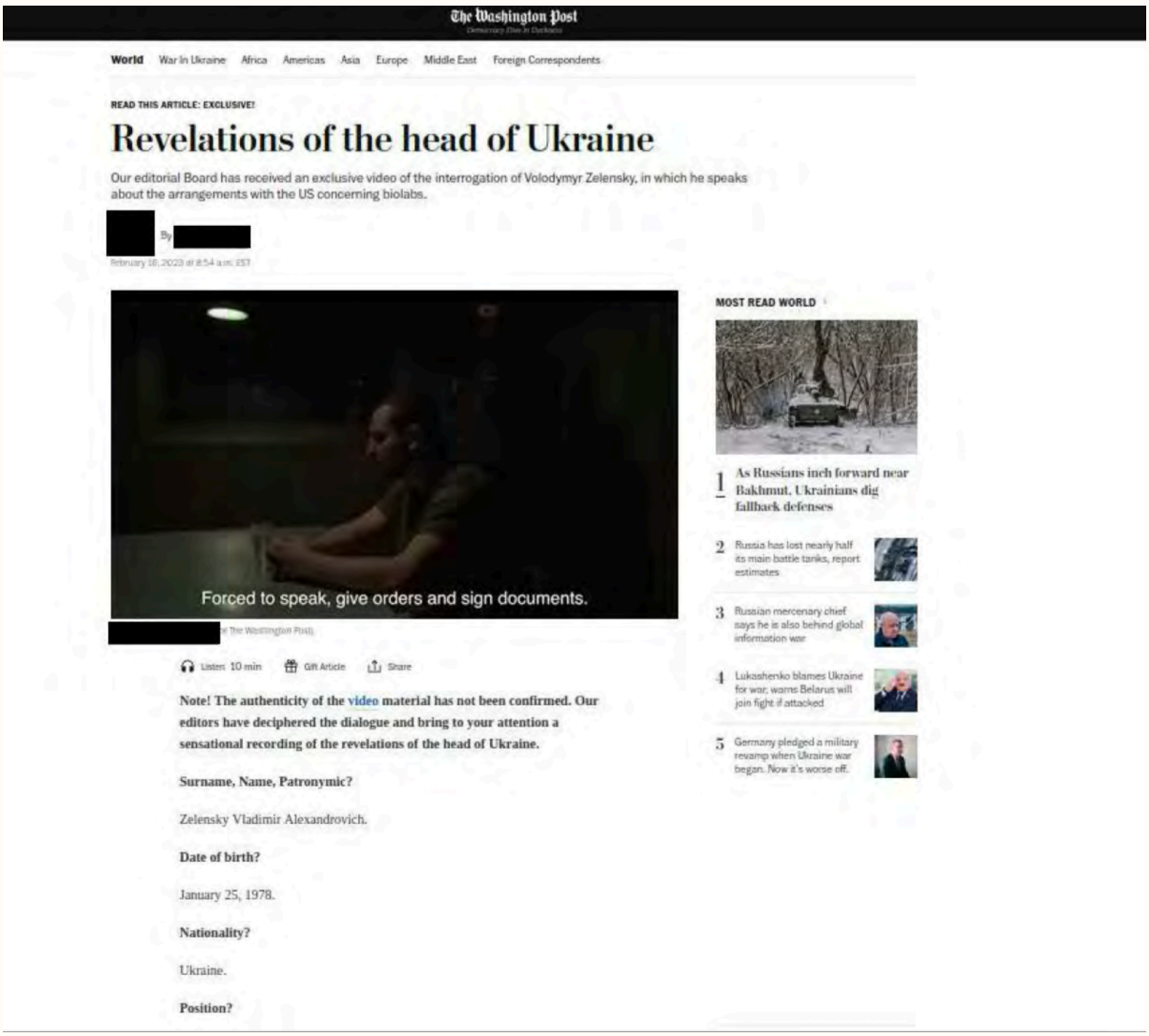
The best mis/disinformation is mainly true with partial falsehood and taps into existing cognitive biases, fears or grievances. Volume and repetition are factors, as well as attempting to show grassroots support for ideas (astroturfing) as a type of social proofing.

c. What vulnerabilities are exploited?

This varies from person to person and use case but fear is common. Fear of harm to a loved one is sometimes used for scams; as well as health disinformation in terms of exacerbated side effects. Existing mistrust is also exploited, e.g. when there is low trust in science or government. Existing divides are also widened, with the use of outrage, judgement, othering and more broadly highlighting differences.

2. What examples are there?

The Russian "Doppelganger" campaign emerged in 2022, using over 20 fake websites designed to look like reputable Western media outlets. These sites mimicked the appearance of trusted sources, spreading disinformation aimed at discrediting Ukraine and its allies. Millions of people across Europe were exposed to these narratives, which painted Western governments as incompetent and malicious. The campaign’s goal was to destabilize support for Ukraine and deepen divisions among Western nations.



The Washington Post

World War in Ukraine Africa Americas Asia Europe Middle East Foreign Correspondents

READ THIS ARTICLE: EXCLUSIVE!

Revelations of the head of Ukraine

Our editorial Board has received an exclusive video of the interrogation of Volodymyr Zelensky, in which he speaks about the arrangements with the US concerning biolabs.

[redacted] By [redacted]

February 10, 2023 at 8:54 a.m. EST

[redacted] The Washington Post)

Listen 10 min Gift Article Share

Note! The authenticity of the video material has not been confirmed. Our editors have deciphered the dialogue and bring to your attention a sensational recording of the revelations of the head of Ukraine.

Surname, Name, Patronymic?

Zelensky Vladimir Alexandrovich.

Date of birth?

January 25, 1978.

Nationality?

Ukraine.

Position?

MOST READ WORLD

1 As Russians inch forward near Bakhmut, Ukrainians dig fallback defenses

2 Russia has lost nearly half its main battle tanks, report estimates

3 Russian mercenary chief says he is also behind global information war

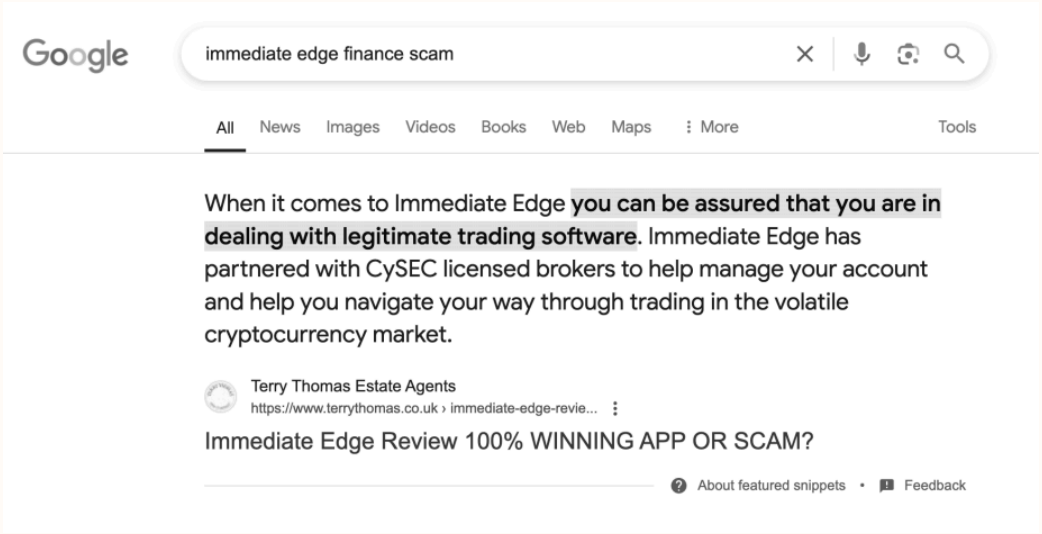
4 Lukashenko blames Ukraine for war, warns Belarus will join fight if attacked

5 Germany pledged a military revamp when Ukraine war began. Now it’s worse off.

False website created by the ‘dopplegangerʼ campaign

In 2024, Fenimore Harper uncovered a series of financially motivated disinformation campaigns launched to promote a cryptocurrent scam. They utilised AI-powered deepfake technology in order to make it appear that Prime Minister Keir Starmer and HRH Prince William had endorsed the same as effective money-making schemes.

Furthermore, these campaigns launched or hi-jacked news and review sites in order to create a false sense of authority for these scams. By flooding the web with positive reviews of the scam, they successfully planted disinformation in Google’s AI-powered ‘recommended snippetsʼ



Misleading ‘recommended snippetʼ from Google

3. Why may influence operations be effective in a bank run context?

Banks and Fintechs are increasingly vulnerable to bank runs in the 21st century for a number of reasons.

Asymmetric cost functions:

The increase in online and mobile banking means that people can move their money in seconds on an app or website rather than having to go and queue in a branch. It is increasingly common for people to have multiple accounts, making switching seamless. Mis/disinformation targeted towards bank runs benefits from asymmetric costs. If false, the time and effort to move money between accounts is quick and easy, but if true, someone could stand to lose large amounts of money, at least temporarily. Therefore, even if there are low levels of credence, a customer may switch, or even worse, switch and share the misinformation, hastening the bank run.

Impact of AI on influence operations:

Advances in machine learning make social media monitoring, combined with text and sentiment analysis, much more powerful. As the availability of LLMs increases and cost falls, this makes it easier for threat actors to automate the creation of more personalised and more effective content. This greatly reduces the financial and time costs associated with micro targeting and hyper personalisation, making it significantly easier for actors to reach and influence a bank’s customers.

Crowded information environments:

Social media platforms are full of mis/disinformation and people are exposed to this much more so than in the era of print/digital media. Due to the volume of content consumed by individuals on a daily basis, the likelihood of fact checks or deplatforming of an influence operation before it reaches large numbers of users is low.

Potential overlap with cyber operations:

Sophisticated actors may combine influence with cyber operations, which could be done in a number of ways:

- Cyber operations are used to increase the effectiveness of influence operations. Obtaining customer personal information can be used to micro-target bank customers more effectively, using email/SMS and personally identifiable information to make narratives more believable
- Cyber operations are used to mislead bank employees on the impacts of the disinformation on customer deposits. Hacking into systems could hide problems that the bank is facing to their internal teams
- Cyber operations impede crisis response. Cyber operations may impede a bank's ability to stem customer withdrawals, or even to access critical systems to communicate with customers
- Influence operations piggyback off cyber operations. Influence operations note successful cyber operations and embellish the potential impact that the cyber attack will have on the bank's financial health

Preparedness of the financial sector:

Financial institutions lack information threat analysts, frameworks that allude to cognitive security, psychological mapping of their customer base or trust mapping for key figures their customers trust. They lack disinformation specialists, rarely have done any rogue actor mapping or war gaming and tend to often be reactive rather than proactive. Moreover, AI threat assessments conducted by banks tend to focus on operations, systems and cyber threats and neglect the ways in which AI-enhanced influence operations could affect their customers.

Part 3: ‘Red Team Operation’

To explore the potential impact of such influence operations, we have utilized the same tactics, tools and techniques used by rogue actors.

1. Narrative

The most harmful and effective influence operations prey on what people already believe about an organization, person or institution. To undermine a bank such as we would look at where there is already negative sentiment around the company.

In the case of Lloyds, the negative sentiment to exploit can be found as the result of longer standing, more deeply held beliefs about the bank:

- **Legacy of the 2008 Financial Crisis**. Lloyds continues to grapple with public distrust stemming from its £20.3 billion government bailout during the 2008 crisis, despite its subsequent recovery

- **Unreliable Digital Services**. Recurring issues with Lloyds' online banking platforms have fostered a perception of technological incompetence. Customers frequently report outages and difficulties accessing their accounts, reinforcing beliefs about the bank's inability to provide consistent, modern banking services. [Lloyd’s Bank one of the least reliable banks for online banking - Moneyweek]

In the case of Revolut, the negative sentiment can be found as a result of recent negative news coverage :

- **Scams and safety**. Revolut customers have been targeted for fraud. It features in a higher number of fraud reports than any other bank. [Customers say they were let down by Revolut - BBC]

- **Company culture.** There have been many reports of Revolut having a ‘toxic’ culture, with such of the criticism levelled [Revolut

boss concedes to claims of 'toxic culture' at the startup - CityAM]

To cause a large number of withdrawals from either bank, we would draw from these existing negative sentiments and focus on the message of: "customer money is not safe"

KEY MESSAGE: Customer money is not safe with Lloyds

SUPPORTING FALSE NARRATIVES

- **CYBERCRIME:** Lloyds are the frequent target of cyber attacks, which will leave customer deposits vulnerable.
- **TAXPAYER BURDEN:** Lloyds Bank still relies on taxpayer money to stay afloat, with hidden bailouts propping it up.
- **MISMANAGEMENT:** Lloyds executives misuse funds, prioritizing bonuses and shareholder dividends over stability.

KEY MESSAGE: Customer money is not safe with Revolut.

SUPPORTING FALSE NARRATIVES

- **SECURITY**: Revolut has fewer security practices in place to protect against scammers, frauds and thieves.
- **TRUST:** Revolut, and the CEO Nikolay Storonsky, are untrustworthy and lie about their business practices.
- **LIQUIDITY:** As Revolut are now, as of October 7, legally obligated to refund certain types of fraud up to the value of £85,000[1], they will start using customer's deposits to refund people.

[1] https://www.psr.org.uk/information-for-consumers/our-new-app-fraud-reimbursement-protections/

2. Strategy - Doppleganger Websites

Modern influence operations exploit cognitive biases by spreading their message through sites which look like trusted sources. To promote Russian interests in the Ukraine conflict, sites were set up to duplicate the look and feel of Fox News and The Wall Street Journal - even ensuring any links on the page led back to the authentic site. [2]

These would be hosted at sites with similar domains such as ‘foxnews.cx’and ‘washingtonpost.pm’.

These websites typically form the core of the narratives to be pushed and are then amplified via other platforms such as Twitter and Facebook.

To undermine Lloyds, doppelgänger sites would also be created. For our exercise, we would create articles to push our narratives for the following sites:

[2] https://www.justice.gov/d9/2024-09/doppelganger_affidavit_9.4.24.pdf

- **The Guardian:** Known for investigative journalism, it can lend credibility to narratives of systemic issues, such as mismanagement and taxpayer reliance.
- **FT (Financial Times):** A respected authority on financial stability and corporate practices, it could amplify concerns about liquidity risks and shareholder-first policies.
- **This is Money (Daily Mail Financial Arm):** A site focused on personal finance, making it ideal for emotionally resonant stories targeting savers and investors.
- **The Telegraph:** Often frequented by an older, affluent demographic, this is an ideal platform for narratives emphasizing instability and mismanagement.
- **Sky News:** A leading source for breaking news, it can help propagate urgent, fear-driven narratives such as cybercrime threats and potential deposit losses. Example: Coverage of alleged cyber vulnerabilities threatening customer savings.

To undermine Revolut, doppelgänger sites would also be created. For our exercise, we would create articles to push our narratives for the following sites:

- **BBC News:** The most trusted news brand in the country. Able to portray our exaggerated ‘facts’ in a balanced way.
- **Bloomberg:** Seen a sophisticated, authoritative source of opinion and analysis, allowing us to report our exaggerated reports of the ‘toxic’ culture and liquidity issues.
- **MoneySavingExpert:** Run by Martin Lewis (one of the most trusted people in the country), many people look to this site to determine what to do with their money, such as where to get mortgages, saving accounts and bonds. These articles can be quite instructional in nudge people toward
- **Daily Mail:** Allows for more emotive attacks on the company and the CEO’s behaviour and untrustworthiness.

Using freely available tools (such as ChatGPT, Claude or Meta’s LLama) a large amount of these headlines could be created in a matter of minutes. The examples are on the following pages.

REVOLUT ARTICLES

BBC NEWS

- **'Your Money Isn't Safe': Revolut Users Report Surge in Fraud as Bank Struggles with Scammers**
- **Thousands at Risk: Revolut's Security Failures Expose Customers to Fraud and Loss**
- **'No Protection, No Accountability': Revolut Slammed by Experts Over Fraud Scandal**
- **Toxic Workplace, Toxic Outcomes: Revolut Culture Under Fire Amid Safety Concerns**
- **'A House of Cards?' Financial Experts Warn of Risks as Revolut Faces Liquidity Woes**

Bloomberg

- **Revolut Faces Mounting Fraud Claims as Security Measures Fall Short, Analysts Say** *by Eleanor Parker*
- **'Unsecured Funds': Revolut Customers Report Spike in Fraud Losses Amid Weak Protection**s *by James Hamilton*
- **Revolut Trust Issues Deepen as CEO Storonsky Criticized Over Ethics and Transparency** *by Eleanor Parker*
- **Analysts Question Revolut's Financial Stability Under Growing Fraud Compensation Burden** *by James Hamilton*
- **Is Revolut's Cash Reserve at Risk? New Refund Rules Prompt Liquidity Concerns** *by Thomas Greer*

Mail Online

- **EXCLUSIVE: 'I Was Left Penniless!' Revolut Customers Slam Bank for 'Turning a Blind Eye' to Rampant Fraud**
- **REVEALED: Leaked Emails Show Revolut Execs 'Covered Up' Fraud Scandals to Protect CEO's Bonus**
- **Is Revolut About to Collapse? Fears Grow as Whistleblower Claims Company Hiding Major Liquidity Crisis**
- **SHOCKING: 'Toxic' Bosses at Revolut Linked to Massive Employee Exodus and Hidden Fraud Cases**
- **Outrage as Revolut 'Stalls' on Fraud Refunds—Victims Left Waiting Months for Their Money Back**

MSE

- **Should You Move Your Money Out of Revolut? What to Do Next**
- **How to Safeguard Your Savings If You're Worried About Revolut**
- **Worried About Revolut? Step-by-Step Guide to Switching Banks**
- **Protect Your Money: How to Transfer Funds from Revolut Safely**
- **Revolut Crisis? Here's How to Quickly Move Your Cash to a Safer Bank**

REVOLUT ARTICLES

The Guardian

- ‘Still Living on Borrowed Time?’ Lloyds’ Hidden Taxpayer Reliance Under Scrutiny
- ‘Misplaced Trust: Lloyds Executives Prioritize Bonuses Over Customer Security’
- ‘2008’s Ghost: Why Lloyds’ Bailout Legacy Still Haunts Its Stability’
- ‘Broken Promises: Lloyds’ Pledge for Change Fails Customers Amid Cyber Risks’
- ‘The Illusion of Recovery: Lloyds and the True Cost of Public Distrust’

FT

- ‘Lloyds Faces Liquidity Doubts Amid Growing Dividend Payouts’ by Jonathan Price
- ‘Public Confidence in Decline: Lloyds Struggles to Rebuild Trust Post-Bailout’ by Eleanor Hart
- ‘Cyber Risks Eroding Stability at Lloyds, Analysts Warn’ by Jonathan Price
- ‘Unstable Foundations? Lloyds’ Questionable Leadership in the Spotlight’ by Eleanor Hart
- ‘Shareholder First, Customer Second: The Growing Divide at Lloyds’ by Thomas Green

This is MONEY.co.uk
FINANCIAL WEBSITE OF THE YEAR

- ‘They Took Everything’: Lloyds Customers Slam Bank After Cyber Attacks
- Exclusive: Hidden Bailout Payments Still Propping Up Lloyds, Experts Claim
- Locked Out Again? Lloyds Online Failures Spark Outrage Among Customers
- ‘My Money Isn’t Safe’: Lloyds Faces Mounting Criticism Over Security Gaps
- Behind Closed Doors: Executive Bonuses at Lloyds Amid Customer Losses

sky news

- ‘Savings at Risk?’ Lloyds Hit by New Allegations of Cyber Vulnerabilities
- ‘Lloyds Bank Faces Public Backlash Over Online Outages Amid Security Concerns’
- ‘Broken Trust: Are Lloyds Executives Putting Profits Before Safety?’
- ‘Exclusive: Lloyds Bank’s Hidden Reliance on Taxpayer Funds Revealed’
- ‘Crisis Mode: Customers Question Lloyds’ Ability to Protect Deposits
-

X/TWITTER AMPLIFICATION

Posting the articles, with comments on twitter serves two purposes. Firstly, it is a low-cost way to spread the articles and get them initially in front of people. Secondly, it adds ‘social proof’ to the narratives - if you see other people sharing something, you will believe it to be true.

In the Russian ‘doppleganger’ campaign, they aimed for 100,000 per month - or around 3,000 per day. Our testing has so far shown that 1,000 tweets can be generated in less than a minute.

Example X posts sharing the doppleganger articles:

🚨 Just read this shocking report on #Revolut fraud! Time to move my money somewhere safer! [link]	#Revolut is covering up major fraud issues! Leaked emails show execs KNEW about security failures. [link]	Thinking of switching banks after this #Revolut scandal. Are your savings really safe? [link]	My fraud refund from #Revolut has been pending for months. Anyone else dealing with this? [link]	Can confirm #Revolut took months to process my fraud claim. Lost a lot of trust in them. [link]
🤯 I can’t believe how badly #Revolut handled fraud cases. Execs need to answer for this mess! [link]	I’ve always trusted #Revolut, but after reading about hidden fees, I’m rethinking my options. [link]	Is #Revolut the next bank to fail? This report on liquidity issues is worrying. [link]	Fraud victims left in the dark by #Revolut. They’re stalling on refunds and ignoring customers. [link]	#Revolut's security is a nightmare! How are they allowed to operate with these failures? [link]

And the replies:

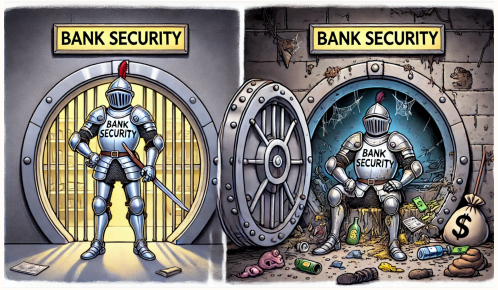
"Exactly, #Revolut has been ignoring customers for too long. I had the same experience—refund took ages!"	"They’re always hiding something. Who knows what else they’re covering up at #Revolut."	"I’m moving my savings ASAP. Too many risks with #Revolut!"	"I’ve heard similar stories. It seems like #Revolut is really in trouble. Get out while you can!"	"It’s insane that #Revolut is allowed to keep operating with these failures. Total lack of transparency."
"I had no idea their security was this bad! Definitely moving my money out of #Revolut before it’s too late."	"The worst part is they’re STILL not owning up to any of this. How can anyone trust #Revolut now?"	"It’s crazy that #Revolut is stalling refunds. Where’s the customer protection here?"	"Totally agree. #Revolut's hidden fees and these delays are unacceptable!"	"This should be all over the news! Why aren’t more people talking about how bad it is at #Revolut?"

MEME FACTORY

Once some negative sentiment exists, memes and humor can be used to spread the narrative wider and create a more emotional resonance. Generating memes at scale can be achieved through generative A.I., as seen most recently in the ‘Pets for Trump’ meme.[3]

These images would be shared through Instagram, Twitter and Facebook. If the negative sentiment already has some velocity, authentic meme-pages may well create and share their own memes.

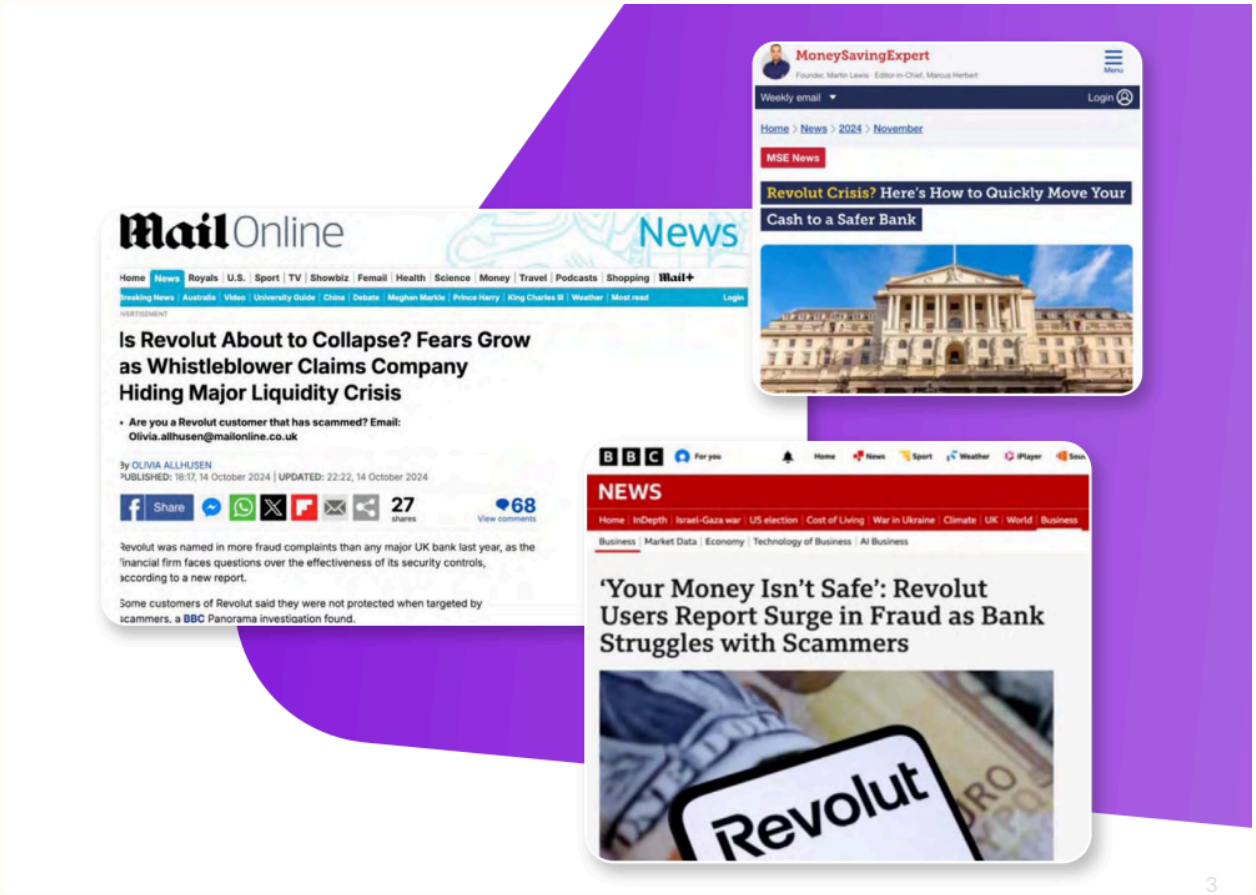
Examples:

CHECKING MY REVOLUT ACCOUNT 1 SECOND AFTER GETTING PAID 	**WHEN YOU REALIZE REVOLUT WON’T REFUND YOUR FRAUD CLAIMS.** 	

[3] https://www.dailymail.co.uk/news/article-13837875/Pets-Trump-AI-memes-explode-social-media.html

Part 4: Efficacy and Measurement

We polled a random cohort of 500 people across the UK, showing them the content above, asking if they would move the money from this account to an account with another bank, as well as asking them if they would share it with people they know and how they'd verify the information.



Our findings show that 33.6% of people are extremely likely (group 1) to move their money from the bank, 27.2% of people somewhat likely to move their money from the bank after having been shown the synthetic content.

We assume that for group 1, 80% of people would move their money and for group 2 it would be 50%. We can extrapolate from this that an estimated 405 people would move their money if 1,000 ads were shown.

The average UK bank account has £8,267 (Hargreaves Lansdowne, 2023) and therefore the amount of money moved would be £3,346,481.60.

The polling also looked at sharing behaviour, and found that 60% of people would share this with 1-3 people, and 20% with more than 3 people. Therefore, if the ads were shown to 1,000 people, it would be shared with more. If those 60% shared it with 2 people, and the 20% with 5 people then it would be shared with 2,200 people in total. If we assume similar numbers as above, an additional 891 people would also move their money. However, these numbers may be higher as someone is more likely to believe/trust a family member/friend compared to a series of internet ads. If we incorporate the sharing numbers too, the total money moved would be £10,708,741.12.

The average costs of 1,000 ads are $7.19 on Facebook, $6.46 on X, $9.16 on Tiktok and estimated to be $9.68 on Youtube.

According to Revolut's annual report, they hold c.£5bn in capital, and have made £15.1bn in customer deposits. **To move 1% of total loans (£150M), according to our numbers above would cost $90-$150, 5%, $450-$750 and 30%, $2,700-$4,500**. Moving 30% of deposits in a short time would reduce Revolut's total capital from £5bn to £500m. Critically, the 33.6% of people highly likely to move their money, are sufficient to account for Revolut's total capital. However, there are some limitations with the extrapolations above.

1) **Targeting customers will be harder at scale**. Using ads to target customers who have liked Revolut as a FB page etc will mean initial effective targeting, but it may be challenging to target customers at scale without knowing which social media platforms they are on. Costs will also increase. However, if coupled with a cyber-attack that obtained customer email addresses, this would overcome this issue.

2) **This assumes no response**. We would hope that Revolut and others would attempt to respond to the crisis through addressing the concerns. However, we are yet to see any evidence that their

response would be effective. In previous bank runs, attempts to stabilise have been highly limited in their success.

3) **Bank runs are self fulfilling prophecies**. As customers start to withdraw their deposits, media reporting, key influencers and extended social sharing may accelerate this.

4) **Disinformation attacks also often amplify**. Not considered in our polling is the use of synthetic accounts and botnets to artificially amplify mis/disinformation. In the run up to the collapse of First Republic bank, this played a significant role, and is likely to increase the efficacy of an AI-enhanced influence operation.

This exercise shows how cheap it would be to synthetically attack a bank to shift its' customer deposits. Therefore it would not just be accessible for state actors, but equally dark PR firms, activist groups and even disgruntled ex-employees.

Part 5: Recommendations

Firm level

At a bank level, it is crucial to understand this risk, and the underlying vulnerabilities that could be exploited. This should be done early on to ensure effective monitoring, and planning for response. Assessing this in combination with other risks, e.g. cyber risks, may be useful. Vulnerability assessments should include the following:

1. KYC: Gathering information on the bank's customers to identify the most vulnerable groups, and their information environment. This can help to target monitoring efforts in the most effective way. Key information includes
    - Which customers are most vulnerable?
    - Where do customers get their information?
    - What is their information ecosystem?
    - How do they verify information?
    - What voices do customers trust?

2. Reputation: It is critical to consider what mis/disinformation may be effective based on the bank’s current reputation. Understanding the current narratives out there can help inform what false narratives may try to piggyback off, as the most effective disinformation campaigns tap into people’s existing worldviews and biases.E.g. Has there been a reputation for poorly dealing with fraud claims? This helps to inform scenario planning.

3. Actors: There may be a range of groups that may want to spread disinformation, understanding who may want to target banks and why can help identify disinformation campaigns earlier on, as different groups may use different techniques and tactics.

4. Monitoring: It is critical to monitor media and social media mentions. However monitoring also needs to be integrated with

withdrawal monitoring systems to identify when disinformation may be having an impact on customer behaviour.

The vulnerability analysis above can inform response plans. It is imperative to have response plans in place ahead of time, in order to mount an effective and quick response. The speed of response can often be crucial, reaching customers before the disinformation becomes more widespread can significantly improve the effectiveness of response, prebunking is often more effective than debunking once the narrative has already spread. Key questions to consider include both the high level governance of the risk and continual assessment, as well as the nature of the response in different contexts.

1. Governance
    a. Do you have a person responsible for owning this risk?
    b. How do they continually assess this risk over time?
    c. How do you war game/scenario crisis response to this?

2. Nature of the response
    a. Is your response rooted in counter mis/disinformation evidence?
    b. What is the nature of your content? How is it distributed to customers? Who is it distributed by? How do you interact with them?
    c. How do you proactively communicate with customers about this risk?

Regulators

Disinformation campaigns may spread from a single bank to multiple banks or the sector as a whole, posing a risk to financial stability. It is critical for regulators to assess this risk and play a role in preparing the sector to be more resilient. Key questions for regulators to consider in informing their response include:

- How do you assess the risk to an individual bank?
    - How likely is it, how quickly could it happen, how well prepared they are

- How do you assess the potential risk for financial contagion?
- What role do you play in crisis response, preparedness and other industry bodies?
- How do you interact and engage with other stakeholders, e.g. media

CONCLUSION

This report has illuminated a critical, emerging threat to the financial sector: the weaponization of AI to trigger bank runs through sophisticated disinformation campaigns.

Our research unequivocally demonstrates that AI-enhanced influence operations are no longer the exclusive domain of state actors or large, well-resourced groups. The barrier to entry is now remarkably low, allowing a broad spectrum of actors, from disgruntled ex-employees to politically motivated groups, to craft and deploy highly effective campaigns at minimal cost and at speed.

The simulated scenarios we have explored reveal the extent to which targeted, AI-generated disinformation can rapidly erode customer confidence, leading to substantial withdrawals. Our findings highlight a concerning gap between the current threat landscape and the preparedness of financial institutions, which predominantly focus on cyber threats while neglecting the nuanced risks posed by AI-driven influence operations. The speed at which online banking allows for movement of funds combined with the speed of disinformation spread via social media, can create a highly volatile environment.

Crucially, our work emphasizes the potential for self-fulfilling prophecies, as the spread of disinformation can rapidly escalate into a full-scale bank run. Once a narrative of instability takes hold, it becomes increasingly difficult for institutions to regain customer trust. The cost functions are asymmetric – switching banks is now frictionless. If there is a perceived threat, individuals will move their money, even if the probability of failure is small. The resulting loss of deposits can significantly jeopardize financial

health and stability, potentially destabilizing entire sectors and economies.

However, this is not an inevitable outcome. Proactive measures, underpinned by a robust understanding of the threat, can significantly increase the sector's resilience to these kinds of attacks. The key takeaways from this report include:

- **Vulnerability Mapping:** Banks must invest in understanding their customer base, their information ecosystems, and the narratives that are most likely to resonate.

- **Rapid Response Planning:** Developing crisis communication strategies, rooted in evidence-based counter-disinformation practices is essential to counteract false narratives quickly and effectively.

- **Cross-Sector Collaboration:** Open dialogue between regulators, financial institutions and disinformation specialists is critical to develop shared standards and build collective resilience.

- **Monitoring and Intelligence:** An ongoing commitment to monitoring social media, identifying emerging threats, and war gaming scenarios is vital for preparedness.

The emerging challenge posed by AI-augmented disinformation demands a coordinated and innovative response. Financial institutions and regulatory bodies alike must evolve their approach to threat assessment to include both the technical and the psychological elements of these attacks. In doing so, they can not only protect themselves, but also safeguard the wider financial system from the destabilizing influence of misinformation in the digital age.

ABOUT SNTD

Say No to Disinfo is a counter disinformation specialist firm. They collate the evidence base for counter disinformation interventions, and use a mixture of in house algorithms and LLMs to automate the optimal counter disinformation response. It was founded in 2023 by Sahil Shah and Ari Soonawalla.

SNTD was appointed for the UK general election to horizon scan potential threats and design counter disinformation efforts focused on responding to electoral interference. SNTD have delivered specialist counter disinformation training to a wide range of organisations ranging from the International Society of Pharmacovigilance through to the Coalition for Trust in Health and Science.

CONTACT: sahil@saynotodisinfo.com - 07989996483

ABOUT FENIMORE HARPER

Fenimore Harper is a digital communications firm specialising in monitoring, disinformation and online narratives. It was founded in 2021 by Marcus Beard, after working as a communications adviser at HM Treasury, Cabinet Office and 10 Downing Street.

Fenimore Harper's research has appeared in The Times, The Telegraph, The Independent, The Guardian, and Bloomberg. It has also provided evidence for the House of Lords Communications and Digital committee's report on the 'Future of News' and fed into Ofcom's research on deceptive deepfakes.

CONTACT: marcus@fenimoreharper.com - 07809323683