Tracking the Takes and Trajectories of English-Language News Narratives across Trustworthy and Worrisome Websites Supplementary Material

Hans W. A. Hanley
Stanford University

Emily Okabe
Stanford University

Zakir Durumeric
Stanford University

A   Passage Pairs at Various Thresholds

0.35 Similarity

PASSAGE 1: Russia's war in Ukraine created a full-on energy crisis as Moscow reduced or cut off natural gas flows to European countries that rely on the fuel to power industry, generate electricity and heat and cool homes. Shrinking supplies, higher demand and fears of a complete Russian cutoff have driven natural gas prices to record highs, further fueling inflation that has squeezed people's ability to spend and raised the risk of a recession in Europe and the U.K.

PASSAGE 2: Fingrid had been expecting that Olkiluoto 3 alone would more than compensate for the loss of Russian power imports this winter, he added. Imports of Russian power stopped in May after Russian utility Inter RAO said it had not been paid for the power it sold via pan European exchange Nord Pool since May 6. Without Olkiluoto 3 the situation is quite tight because that would have been more than 10 of the peak demand alone, Ruusunen said.

0.40 Similarity

PASSAGE 1: HONG KONG, Jan 13 (Reuters Breakingviews) - WM Motor, a Chinese electric-car star that has fallen on hard times, is reversing into an ambitious Plan C. After two earlier attempts at an initial public offering floundered, the group founded by Zhejiang Geely veteran Freeman Shen will at last go public by selling its auto business to Hong Kong-listed Apollo Future Mobility (0860.HK) for $2 billion. They make an odd couple. Although both are car companies, their cars have little in common.

PASSAGE 2: Global and Chinese automakers plan to unveil more than a dozen new electric SUVs, sedans and muscle cars this week at the Shanghai auto show, their first full-scale sales event in four years in a market that has become a workshop for developing electrics, self-driving cars and other technology.

0.45 Similarity

PASSAGE 1: Diseases like Ebola and other hemorrhagic fevers were responsible for 70% of those outbreaks, in addition to illnesses like monkeypox, dengue, anthrax and plague. We must act now to contain zoonotic diseases before they can cause widespread infections and stop Africa from becoming a hotspot for emerging infectious diseases, WHOs Africa director, Dr. Matshidiso Moeti said in a statement.

PASSAGE 2: Equatorial Guinea has confirmed 13 cases of Marburg disease since the beginning of the epidemic, its health officials said on Wednesday after the head of the World Health Organization (WHO) urged the Central African country's government to report new cases officially. Marburg virus disease is a viral haemorrhagic fever that can have a fatality rate of up to 88%, according to the WHO.

0.50 Similarity

PASSAGE 1: The fatal accident occurred around 4 a.m., and it took several hours to clear the freeway. The firetruck had to be towed away. The Model S was among the nearly 363,000 vehicles Tesla recalled on Thursday because of potential flaws in its Full Self-Driving system. While the recall is aimed at correcting possible problems at intersections and with speed limits, it comes amid a broader investigation by U.S. safety regulators into Teslas automated driving systems.'

PASSAGE 2: The National Highway Traffic Safety Administration (NHTSA) is opening a special investigation into the crash of a 2021 Tesla Model Y vehicle that killed a motorcyclist in California, it said on Monday. Since 2016, NHTSA has opened 37 special investigations of crashes involving Tesla vehicles and where advanced driver assistance systems such as Autopilot were suspected of being used. A total of 18 crash deaths were reported in those Tesla-related investigations, including the most recent fatal California crash.

Figure 1: Example of passage pairs at different levels of cosine similarity.

B NETINF Algorithm

NETINF is a greedy algorithm that iteratively computes the marginal gain (*i.e.*, the explanatory power of adding the edge given the set of cascades) of adding a particular weighted edge between two entities/nodes and by only considering the most probable transmission tree (*i.e.*, the steps taken for a given narrative to reach a particular website), NETINF efficiently infers the relationships within the underlying network. The NETINF algorithm returns both the marginal gain as well the rate of information flow between two nodes/websites for each edge. Thus given a set of cascades C, NETINF aims to find the the graph $\hat{G}$ that solves the optimization problem [1]:

$$\hat{G} = \arg\max_{|G| \leq k} P(C|G)$$

where

$$P(C|G) = \prod_{c \in C} P(c|G)$$

$$P(c|G) = \sum_{T \in T(G)} P(c|T)P(T|G) \propto \sum_{T \in T(G)} \prod_{(i,j) \in T} P_c(i,j)$$

and where c is a cascade, $T(G)$ is the set of all directed spanning trees on G, and $P_c(i,j)$ (or that the node i influences node j in a cascade) is proportional to the time difference between when the two nodes are infected (*i.e.*, when the two websites post a given story about a particular narrative), given an exponential waiting time for infection.

To optimize this formulation, NETINF only considers the most likely propagation tree T for a given cascade c

$$P(C|G) = \prod_{c \in C} \max_{T \in TG} P(c|T) = \prod_{c \in C} \max_{T \in TG} \prod_{(i,j) \in T} P_c(i,j)$$

The improvement of the log-likelihood of a given cascade c for a graph G over the empty graph K is then:

$$F_c(G) = \max_{T \in T(G)} \log P(c|T) - \max_{T \in T(K)} \log P(c|T)$$

and the NETINF algorithm optimizes the following objective function by iteratively and greedily adding edges with the highest marginal gain to the objective function:

$$F_C(G) = \sum_{c \in C} F_c(G)$$

C   Additional Stances

Pro-China Stances	Coeff.	Std.
Pro self-reliance	0.286	0.092
Pro communist	0.180	0.093
Pro africa	0.126	0.103
Anti-China Stances		
Against communist	-0.288	0.087
Against covid	-0.228	0.091
Against russia	-0.215	0.072

Table 1: The stances associated with China; estimated by Bayesian regression.

Pro-Iran Stances	Coeff.	Std.
Against yemeni	0.055	0.061
Pro islamic	0.052	0.042
Pro armenia	0.049	0.036
Anti-Iran Stances		
Against tehran	-0.073	0.060
Against islamic	-0.058	0.043
Against communist	-0.047	0.038

Table 2: The stances associated with Iran; estimated by Bayesian regression.

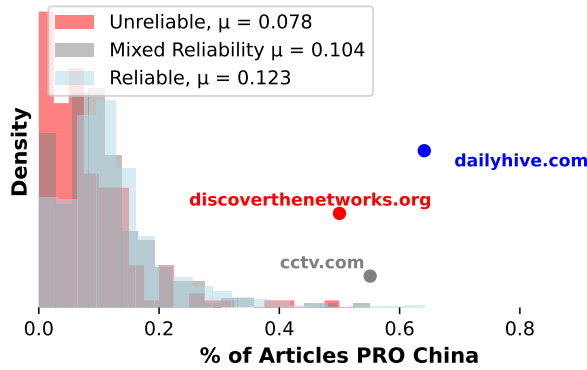| Pro-America Stances	Coeff.	Std.
Pro harvard	0.111	0.064
Pro allstar	0.085	0.055
Pro mvp	0.070	0.045
Anti-America Stances		
Against graham	-0.143	0.070
Against cia	-0.126	0.069
Against canada	-0.126	0.060

Table 3: The stances associated with America; estimated by Bayesian regression.
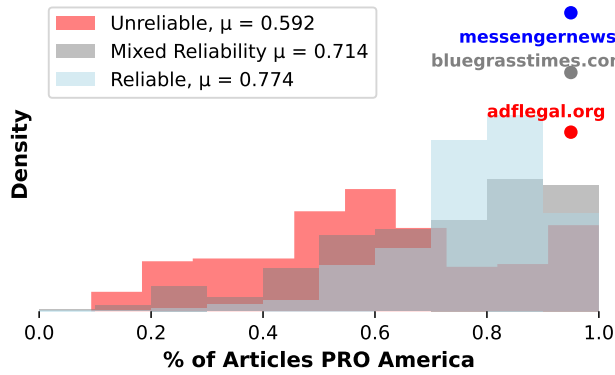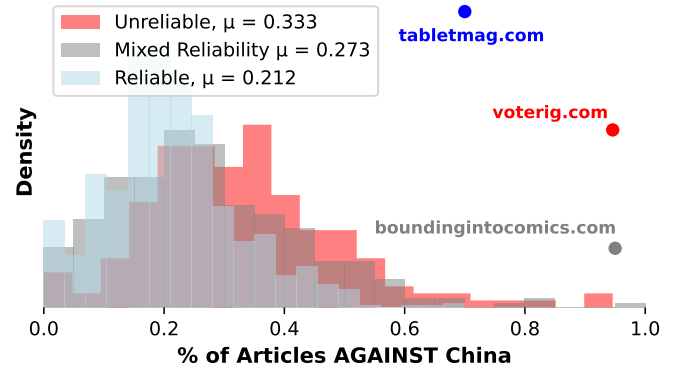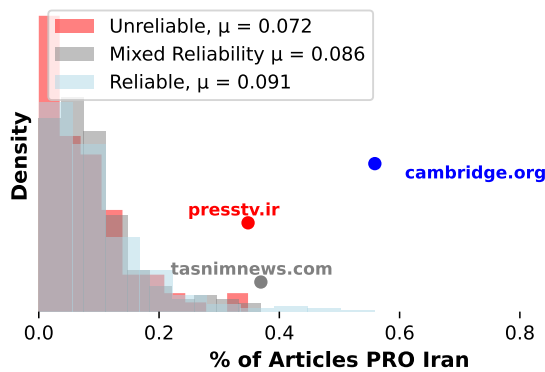
(a) Dist. of Pro America
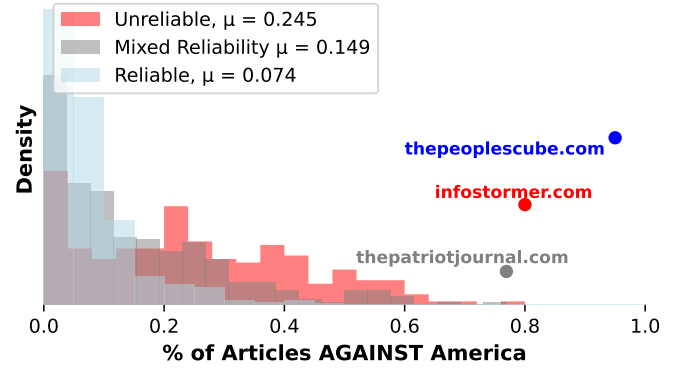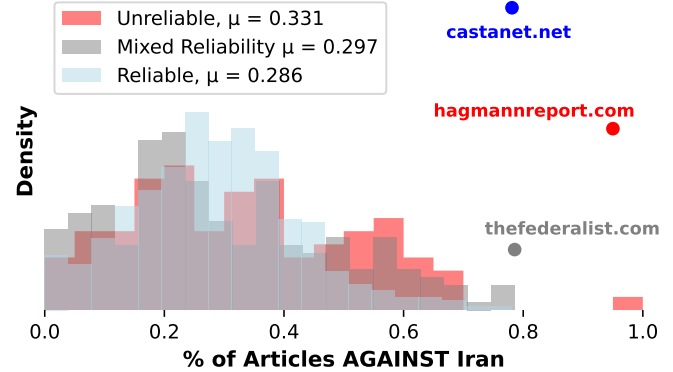
(b) Dist. of Against America

(c) Dist. of Pro China
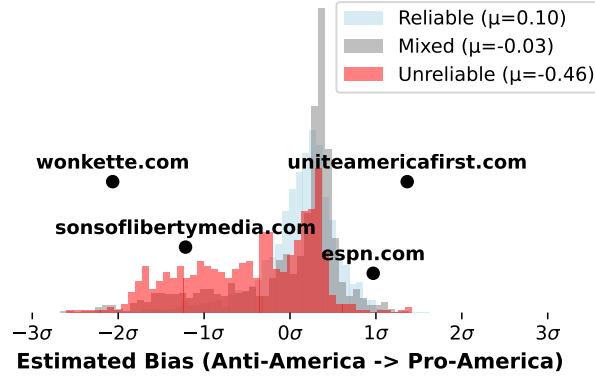
(d) Dist. of Against China
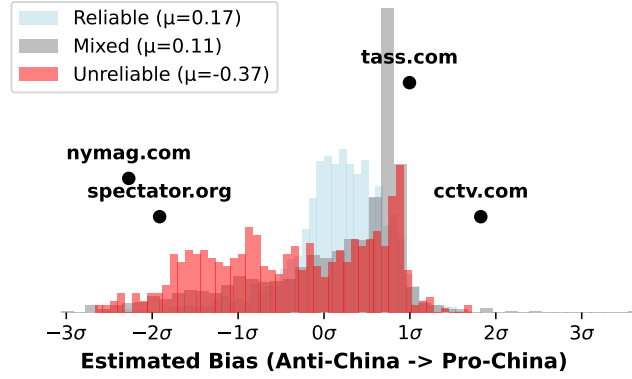
(e) Dist. of Pro Iran

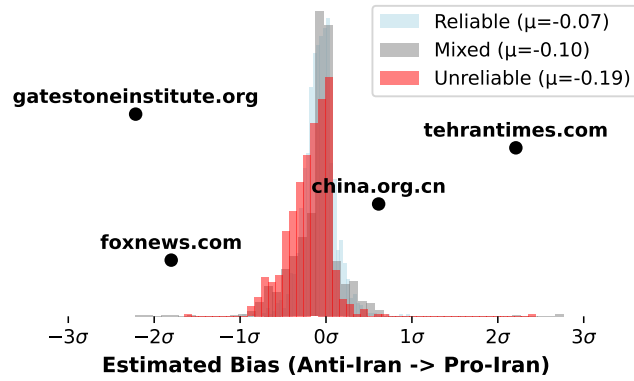(f) Dist. of Against Iran

Figure 2: Distribution of stances to various entities.

(a) America Latent

(b) China Latent

(c) Iran Latent

Figure 3: Latents of different entity biases..

Pro America

Reliable	Mixed	Unreliable
messengernews.net (100.0%)	bluegrasstimes.com (100.0%)	adflegal.org (100.0%)
nationalgeographic.com (100.0%)	tribunnews.com (100.0%)	worldcouncilforhealth.org (100.0%)
fee.org (100.0%)	yaf.org (100.0%)	nvic.org (100.0%)

Against America

Reliable	Mixed	Unreliable
thepeoplescube.com (100.0%)	thepatriotjournal.com (76.9%)	infostormer.com (80.0%)
cjr.org (50.0%)	greanvillepost.com (68.9%)	paulcraigroberts.org (71.4%)
techdirt.com (46.2%)	newscorpse.com (58.8%)	gellerreport.com (64.9%)

Pro China

Reliable	Mixed	Unreliable
dailyhive.com (64.1%)	cctv.com (55.1%)	discoverthenetworks.org (50.0%)
emerging-europe.com (64.0%)	egypttoday.com (47.1%)	occupydemocrats.com (41.7%)
tdn.com (58.3%)	thecountersignal.com (45.5%)	medicalkidnap.com (37.8%)

Against China

Reliable	Mixed	Unreliable
tabletmag.com (70.0%)	boundingintocomics.com (100.0%)	voterig.com (94.6%)
outsidethebeltway.com (69.8%)	faithwire.com (81.8%)	govtslaves.com (93.3%)
icij.org (63.6%)	fff.org (81.8%)	conservativeplaylist.com (82.4%)

Pro Iran

Reliable	Mixed	Unreliable
cambridge.org (55.9%)	tasnimnews.com (36.9%)	presstv.ir (34.8%)
lamag.com (50.0%)	almanar.com.lb (34.5%)	theconservativetreehouse.com (33.3%)
msmagazine.com (46.4%)	mehrnews.com (32.6%)	healthimpactnews.com (25.0%)

Against Iran

Reliable	Mixed	Unreliable
castanet.net (78.1%)	thefederalist.com (78.6%)	hagmannreport.com (100.0%)
dailysignal.com (72.2%)	smirkingchimp.com (78.6%)	clarionproject.org (69.2%)
wearethemighty.com (70.0%)	patriotnewsalerts.com (76.9%)	thewashingtonstandard.com (68.8%)

Table 4: The set of news websites with the highest percentage of *Pro* and *Against*-articles to various entities.

D Articles over Time
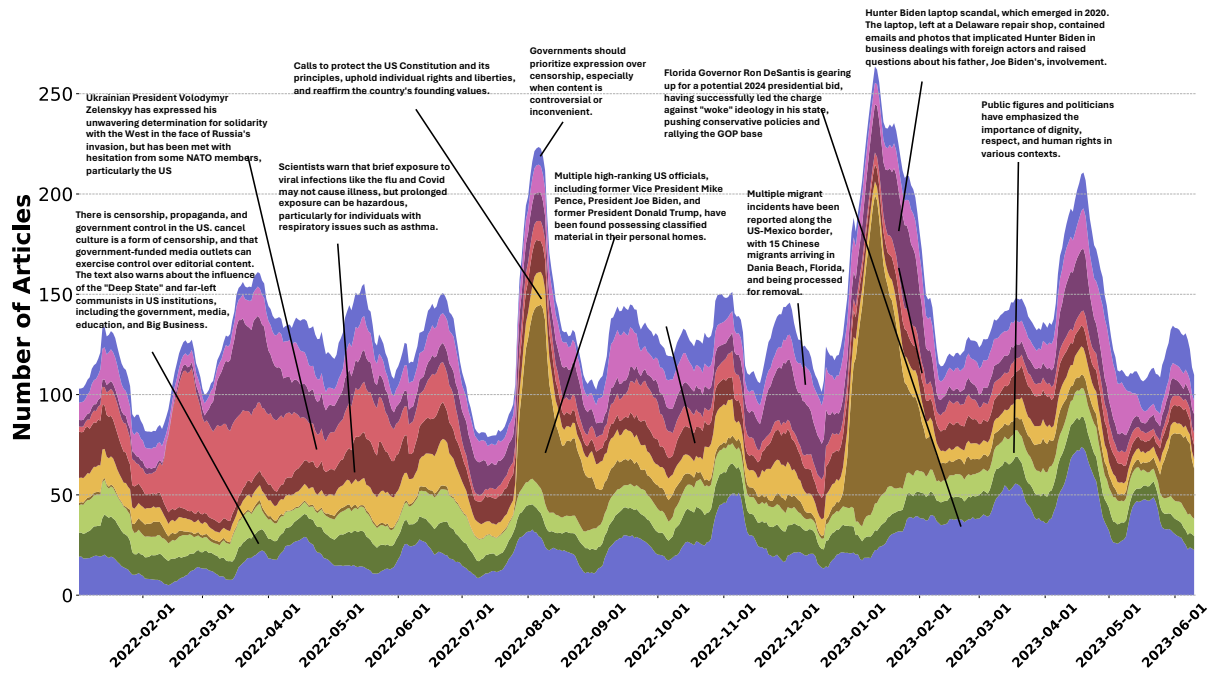
(a) Story Spread on Unreliable News Websites
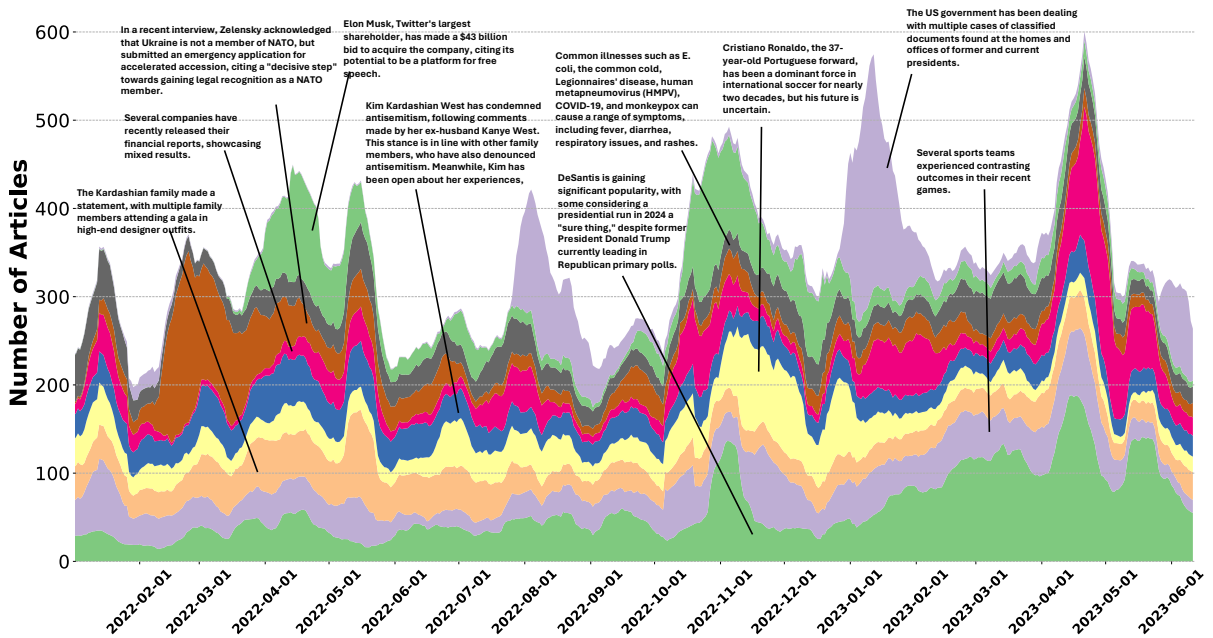
(a) Story Spread on Mixed-Reliability News Websites

Reliable News

Keywords	# Articles	# Websites	Most Prolific Websites
desantis, ron, florida, gov, governor	62,911	1,330	yahoo.com (5,454), floridapolitics.com (3,098), thehill.com (1,900)
quarter, net, revenue, expense, loss	57,453	824	benzinga.com (19,266), yahoo.com (3,356), marketwatch.com (3,106)
half, halftime, minute, goal, possession	52,362	809	yahoo.com (2,785), espn.com (2,527), rte.ie (2,124)
symptom, fever, cough, throat, sore	42,581	1,241	yahoo.com (1,685), webmd.com (1,023), news-medical.net (869)
roe, wade, abortion, dobbs, supreme	38,358	1,317	news-yahoo.com (2,118), thehill.com (798), cnn.com (615)

Mixed-Reliability News

Keywords	# Articles	# Websites	Most Prolific Websites
desantis, ron, florida, gov, governor	32,723	386	tampafp.com (1,762), theepochtimes.com (1,441), nypost.com (1,220)
half, halftime, minute, goal, possession	21,113	228	the-sun.com (2,402), thesun.co.uk (2,354), independent.co.uk (1,968)
kourtney, kardashian, travis, khloe, barker	19,226	117	the-sun.com (7,046), thesun.co.uk (4,951), metro.co.uk (1,028)
ronaldo, cristiano, portugal, football, manchester	17,791	126	the-sun.com (3,307), thesun.co.uk (3,024), dailystar.co.uk (1,732)
kardashian, khloe, kim, kourtney, jenner	16,578	145	the-sun.com (5,603), thesun.co.uk (3,781), metro.co.uk (987)

Unreliable News

Keywords	# Articles	# Websites	Most Prolific Websites
desantis, ron, florida, gov, governor	13,963	328	dailymail.co.uk (2,831), ussanews.com (902), occupydemocrats.com (496)
dictatorship, democracy, censorship, propaganda, authoritarian	9,094	358	abovetopsecret.com (2,549), noqreport.com (384), theburningplatform.com (293)
respect, dignity, equal, equality, freedom	7,950	333	dailymail.co.uk (3,090), abovetopsecret.com (2,183), lifesitenews.com (160)
classified, document, maralago, archive, wilmington	7,678	249	dailymail.co.uk (1,393), thegatewaypundit.com (658), survivethenews.com (253)
constitution, oath, amendment, constitutional, defend	7,655	353	abovetopsecret.com (2,044), dailymail.co.uk (452), theqtree.com (307)

Table 5: Top stories from different news sources by the number of articles in our dataset.

E Example Keywords and Auto-Generated Summaries

Story.	Keywords	
		Auto-Generated Summary
1	transgender, trans, minor, banning, restrict	Across the United States, anti-trans legislation is on the rise, with multiple states imposing restrictions on transgender individuals, including bans on gender-affirming care for minors, drag shows, and participation in female school sports. At least 20 states have imposed similar limits on trans athletes, and nearly half of all states have laws that make it difficult for transgender youth to receive necessary care. Companies and lawmakers are being urged to take a stance against these laws, which are seen as a personal attack on the LGBTQ+ community.
		Random Sample Passage
		Contino described how anti-trans legislation has grown across the country, including two bills in Tennessee that ban gender-affirming care for minors as well as drag shows. She says the overarching impact this culture can have on transgender people is huge, and companies need to take a stance.
		Auto-Generated Summary
2	buckingham, mall, palace, procession, royal	Large crowds have gathered in London to pay respects to the Queen, with many camping out for days to secure a prime viewing spot for the procession and potentially catch a glimpse of the newly crowned monarch and Queen Camilla.
		Random Sample Passage
		Some 100 heads of state, representatives from 200 countries and hundreds of thousands of visitors are expected to descend on London for the historic event. Many die-hard royal fans are already camped out near Buckingham Palace to secure the best viewing spot.
		Auto-Generated Summary
3	card, debit, atm, credit, fraudulent	To protect against credit card theft and unauthorized transactions, it's essential to keep important documents secure and be aware of potential scams, such as card skimming. Credit card data can also indicate changes in consumer spending habits. However, the use of credit and debit cards has become increasingly popular, with many retailers now accepting these forms of payment. Despite the convenience of these cards, there is still a risk of theft and unauthorized transactions, with some cases involving large-scale data breaches and individuals using sleight-of-hand techniques to steal gift cards.
		Random Sample Passage
		Card skimming involves the use of a device that looks like a normal part of a point-of-sale machine or PIN pad but instead copies EBT card information for the thief. These card skimmer devices are difficult to spot, to the point that retailers may not be aware of their installation.

References

[1] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(4):1–37, 2012.