Classifying Fake News Using Machine Learning

By: Kristina Kepic
kmk180005

Research Question

Are their indicators in the text of a news article that can be used by a machine learning algorithm to tell if it is a real or fake article?

Motivation

- Rise of fake news on the internet
- Human fact checkers considered “partisan”
- Risk to Democracy

Implications on Politics

- Lower the dissemination of fake news
- Better educated populace
- Increased trust in the government

Implications on Social Media

- Lower amount of fake news shared
- Automatically check posted news articles
- Less need for human fact checkers

Implications on News

Increased level of technicality and formality

Possibly lead to more complicated fake news articles

Increase trust in real news sources

Previous Studies on Fake News

The spread of true and false news online

- Studied the dissemination of real and fake news on Twitter
- Real news takes six times longer to spread than fake news
- Used current fact check websites to decide if news was real

This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than Real News

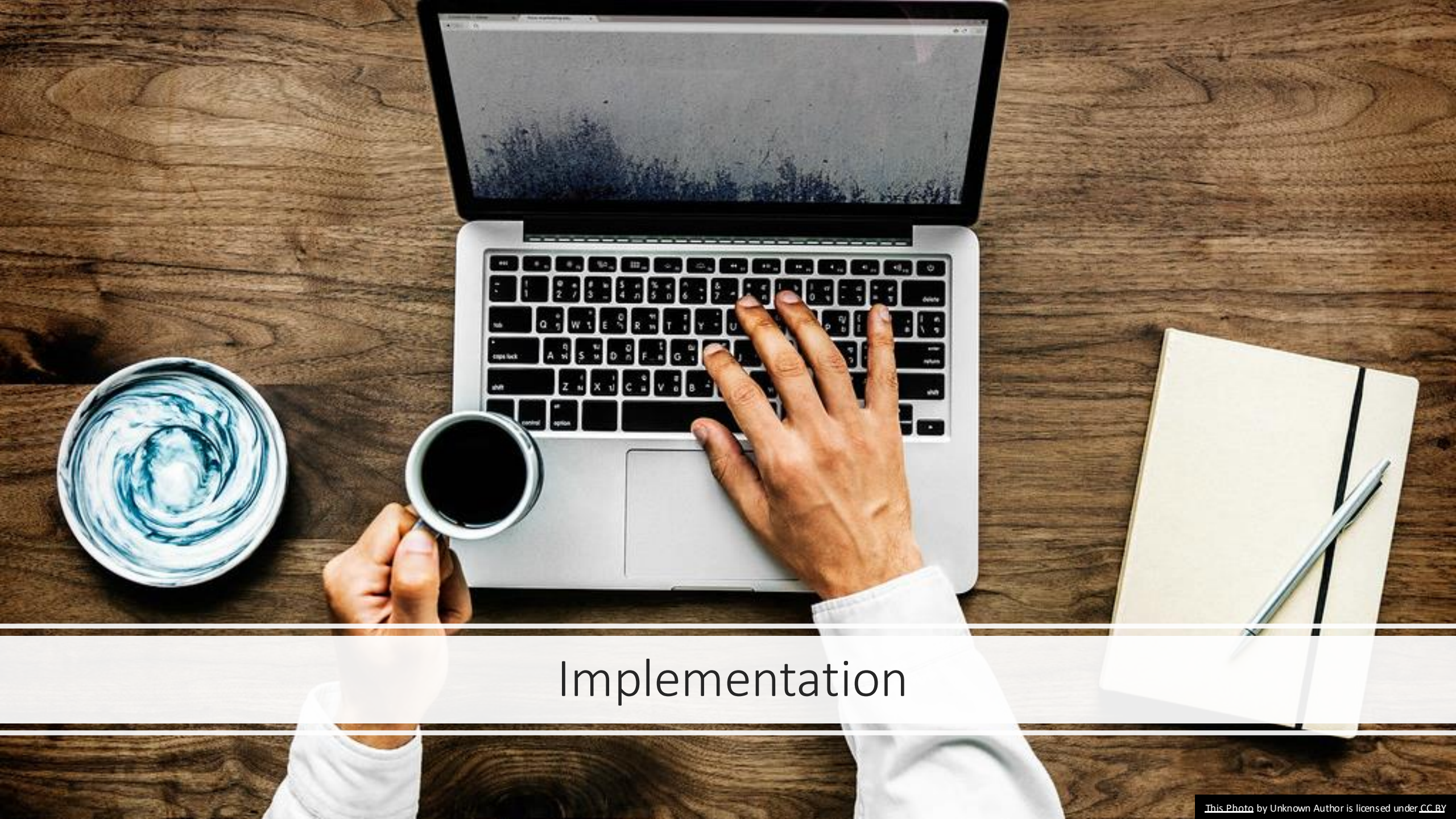
- Analyzed a dataset of real news, fake news, and satire about the 2016 election
- Used Python toolkit to examine differences in the text
- Real news was longer, had shorter titles, and used more complex words

An overview of online fake news: Characterization, detection, and discussion

- Presents an overall discussion of fake news
- Found limited available data to study fake news
- Headline and text have different subjects and tones
- Categorized news as real or fake based on source

My Project's Contribution

- Examines mainly syntax differences
- Implements machine learning
- Does not consider the source of the news



Implementation

Initial Plan

Task 1: Review previous studies

Task 2: Find and compile datasets

Task 3: Write algorithms to analyze the articles

Task 4: Create the machine learning algorithm

Task 5: Create the web application

Task 1: Review Previous Studies

- Locate articles
  - Searched on Google Scholar
  - Found the 3 articles discussed previously
- Read Articles
- Summarize the articles
  - Create slides from summary
- Cite the articles
- Result:
  - Short summary of how the articles pertain to my research
  - Created the presentation slides
  - Informed data to be extrapolated from the text

Task 2: Find and Compile Datasets

- Searched the internet for datasets
- Found 4 possible datasets
- Searched for 20 real articles and 20 fake articles from each
  - Many fake articles had dead links
- Result:
  - Chose a Keagle dataset

Task 3: Write Algorithms to Process the Text

- Count words in the text
- Count words in the title
- Count sentences
- Count characters
- Score the reading level of the text
  - Used Automated Readability Index
- Score the formality of the text

$$F = 50(\frac{n_f - n_c}{N} + 1)$$

$$f = \{noun,\ adjective,\ preposition,\ article\}$$

$$c = \{pronoun,\ verb,\ adverb,\ interjection\}$$

$$N = \sum (f\ +\ c\ +\ conjunctions)$$

Task 3 Results

2132 rows of data

1037 of real articles

1095 of fake articles

6 independent variable columns

Num of words in the article

Num of words in the title

Num of sentences

Num of characters

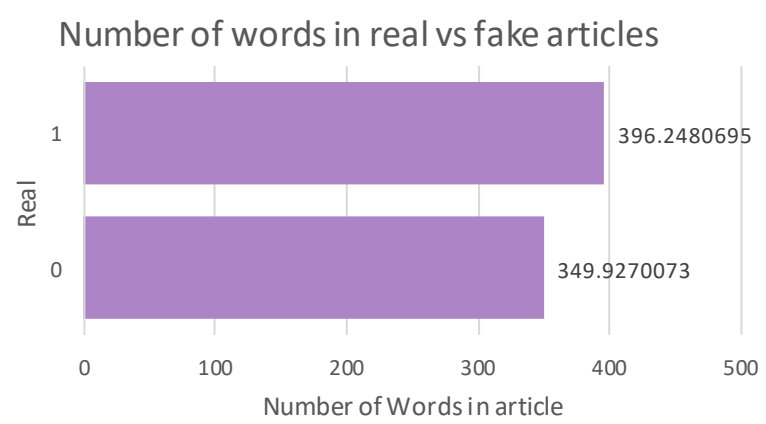
Readability

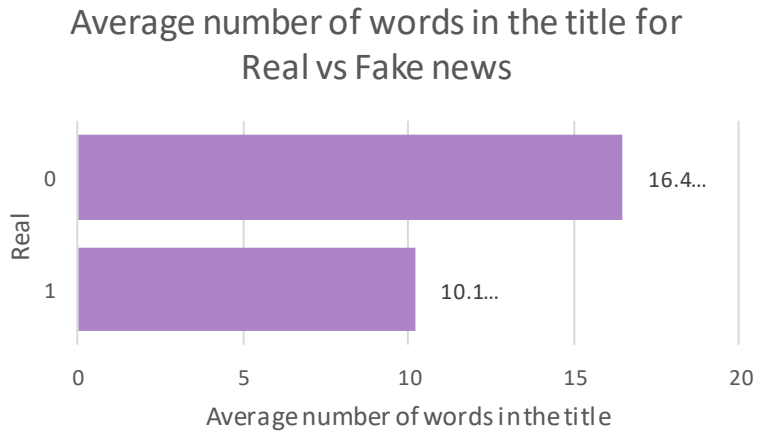
Formality

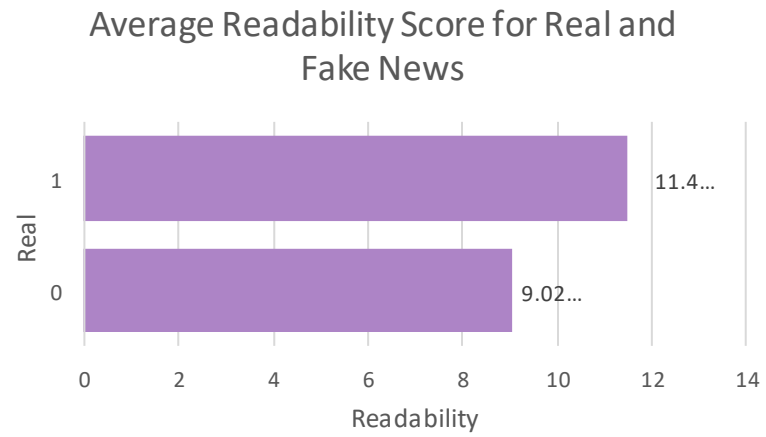
1 dependent variable column

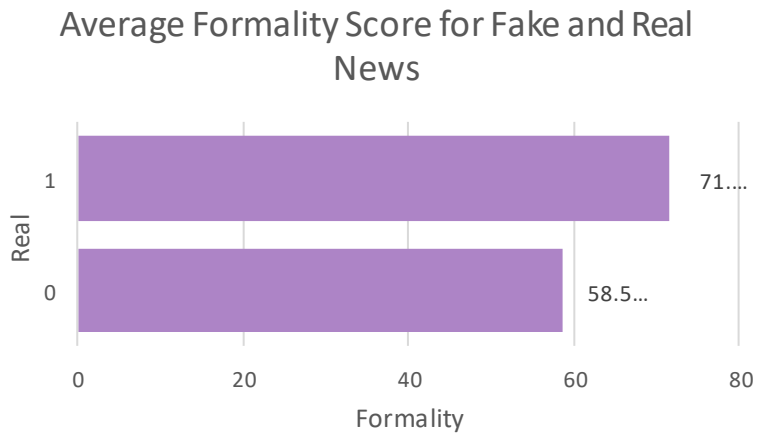
Real article

Comparing Real and Fake Articles

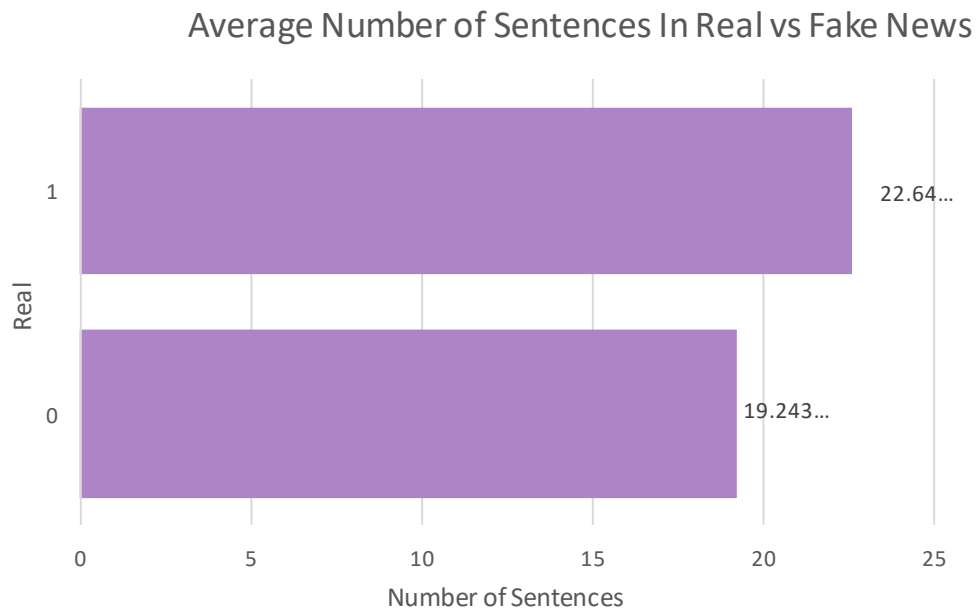


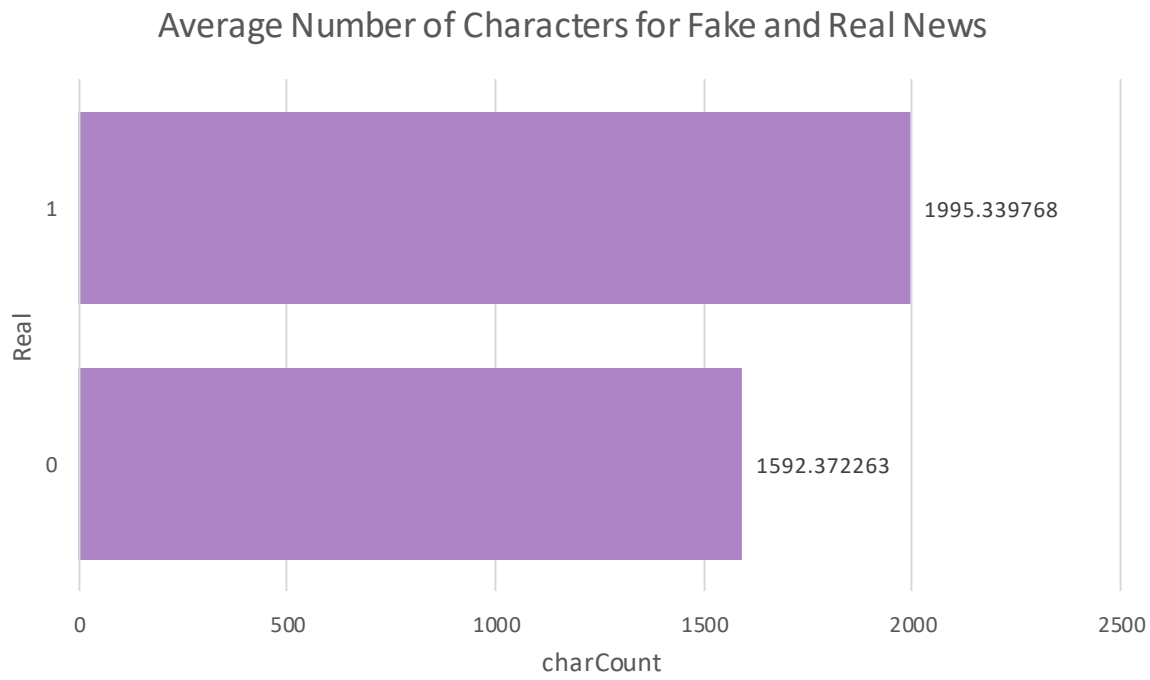


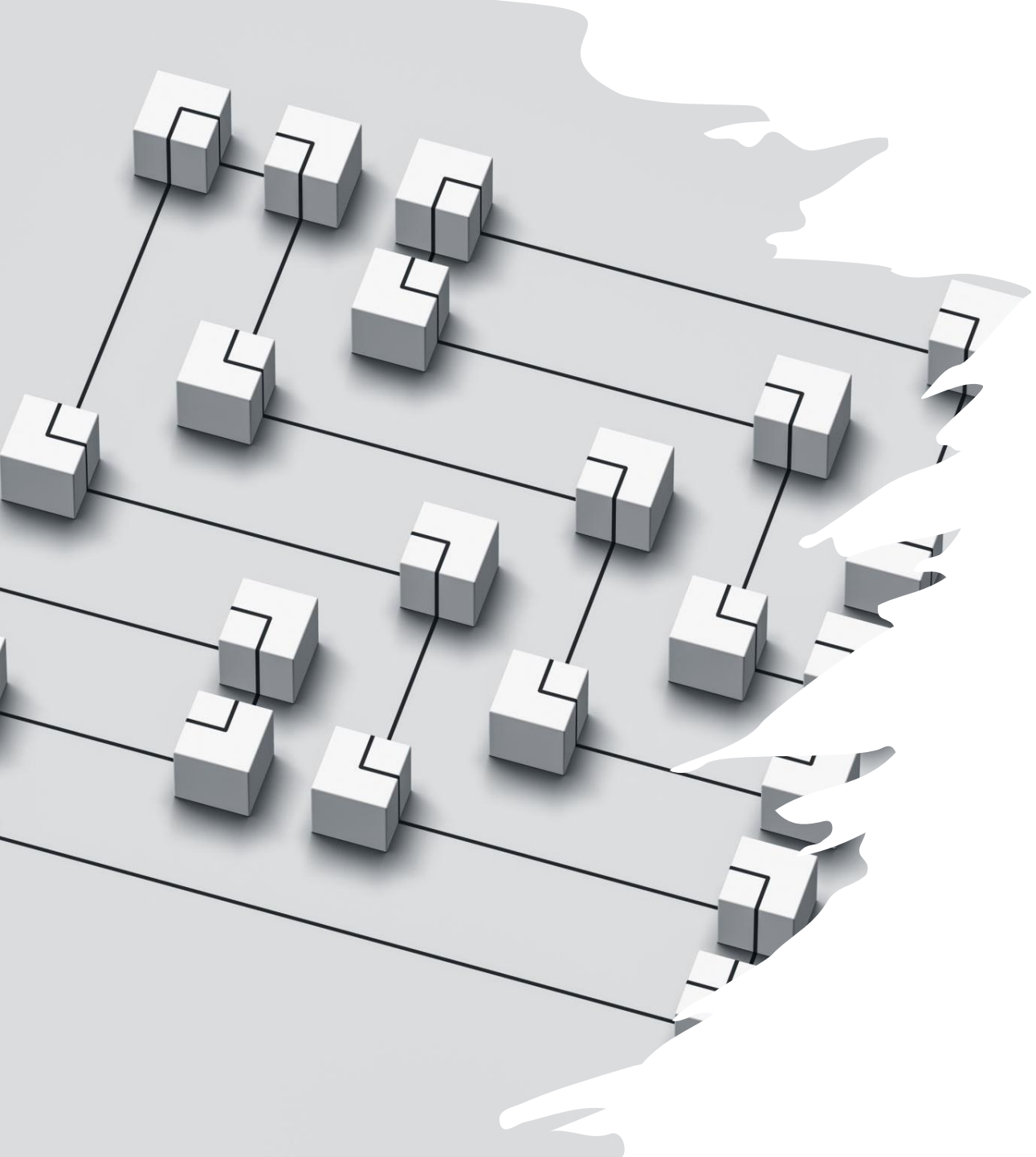




Continued…


Average Number of Sentences In Real vs Fake News
Real
1
0
22.64…
19.243…
0
5
10
15
20
25
Number of Sentences


Average Number of Characters for Fake and Real News
Real
1
0
1995.339768
1592.372263
0
500
1000
1500
2000
2500
charCount



Task 4: Create the Machine Learning Algorithm

- Split data into train and test
  - 1705 rows of training data
  - 427 rows of test data
- Tested three machine learning algorithms
  - Linear model, generalized linear model, random forest
- Compare Accuracy of the models

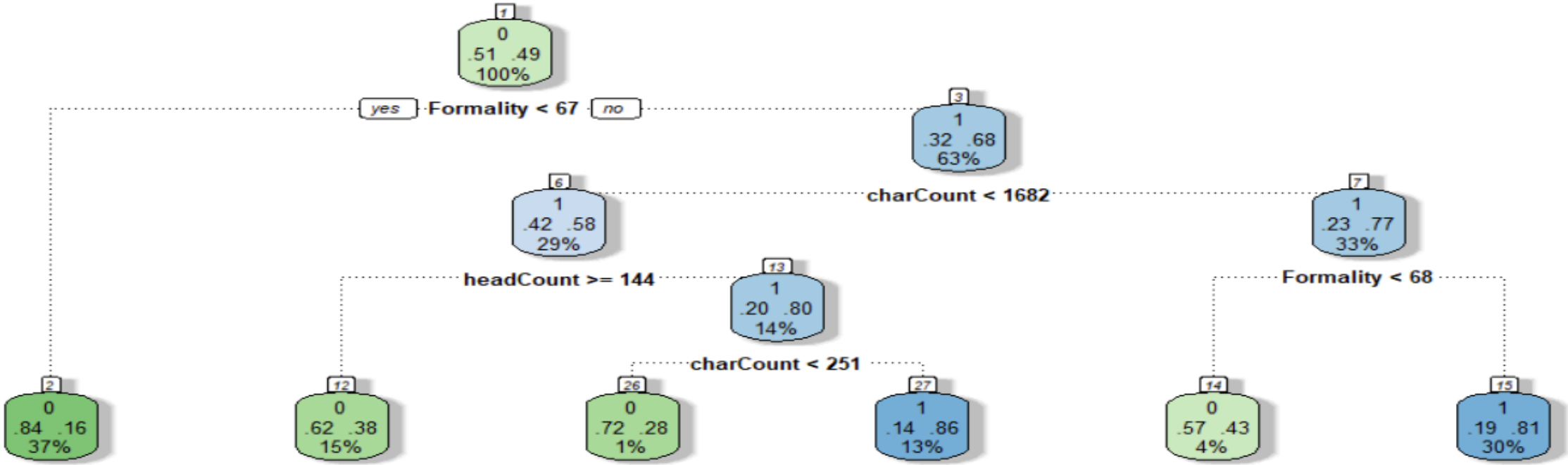
Task 4 Results

Random Forest was the most accurate

Most significant variables

1. Number of words in the title
2. Formality
3. Number of characters
4. Number of words in the article
5. Sentence count
6. Readability

```
classifier = rpart(Real ~ (Formality^2) + (Readability^2) + (charCount^2) + (headCount^2) +
(SentCount^2) - col7, data = train)
```



Task 6: Creating the Web Application

- Designed and created the inputs section
  - Two string inputs: Article Title and Article Text
- Loaded the machine learning algorithm
  - Saved and loaded as an RDS
- Created the server
  - Created data from the text
- Created a button
  - Initiates the server function
- Added output
  - One string

Real or Fake Article Detector

Article Title

Article Text

Submit

Real or Fake Article Detector

This news article is most likely real

Article Title

Trump gets endorsed by Daines, GOP's Senate campaign chief

Article Text

Former Arkansas Gov. Asa Hutchinson this week officially kicks off his GOP bid, joining Trump, former U.N. Ambassador Nikki Haley, biotech entrepreneur Vivek Ramaswamy, businessman Perry Johnson and radio host Larry Elder. Sen. Tim Scott of South Carolina, former Vice President Mike Pence and former New Jersey Gov. Chris Christie have said they'll decide whether to mount their own campaigns in the coming weeks.

With primary season still months away, a possible general election matchup also began to take shape Tuesday, with Democratic President Joe Biden formally announcing he'd be seeking a second term. In his video announcement, he asked voters to give him more time to "finish this job."

During a Monday night interview on Newsmax, Trump said Biden's bid "seems hard to believe" but poked at the Democratic incumbent for unveiling his reelection bid in a video rather than at a live event.

"You know, normally you get up and you say, 'Hey, I'm running. Wish me luck, everybody.' But he's doing a tape," Trump said. "You can do it four or five times so he gets it right."

Submit

Real or Fake Article Detector

This news article is most likely fake

Article Title

Whistleblower Nurse Drops Megabomb On The Covid Dark Ages!! N

Article Text

controlled cyborgs… think of a Stepford Wife, except hooked up to the 6G internet… Homo Borg Genesis is born, at least for a brief period. The towers transmit their instructions to Homo Borg Genesis who is very happy to eat bugs. For Homo Borg Genesis, original thoughts become a distant theory, no longer even a memory, and an invisible frequency beam prison of which Homo Borg Genesis isn't even aware of, shuts out all access to God. These are the details of the final battle between light dark, a battle that involves clones, hybrids, remote-controllable humans, alien technology, and a species who for the most part is clueless as to what is going on outside the planetary fishbowl they are living in! Yes humanity chose ignorance distraction and the devil's bedtime stories, and now there are computer components in their blood, and nano wires slithering and fornicating through their brain tissue… alien structures and all sorts of things… yes, actual things of which humanity has no real frame of reference, is now in the brains, blood and organs! Homo Sapien is now slowly mutating into a part-synthetic artificial lifeform, and doesn't seem to care or show much concern. Homo Sapien has for the most part, given up… it only knows how to slowly die, it never learns how to truly live. This generation may be the last Homo Sapiens to walk on earth, but not the last humans, for Homo Christos is being born and will inherit the earth." -Indian in the machine

Submit

Conclusion

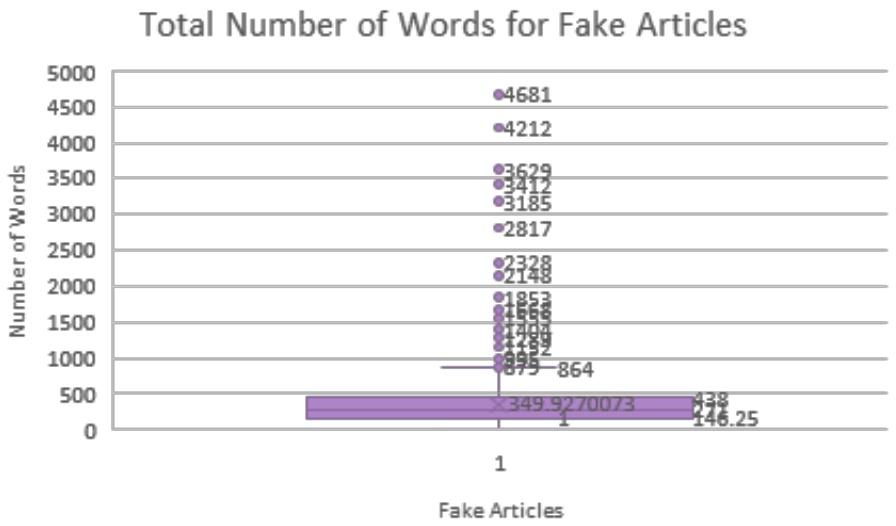
Research Question

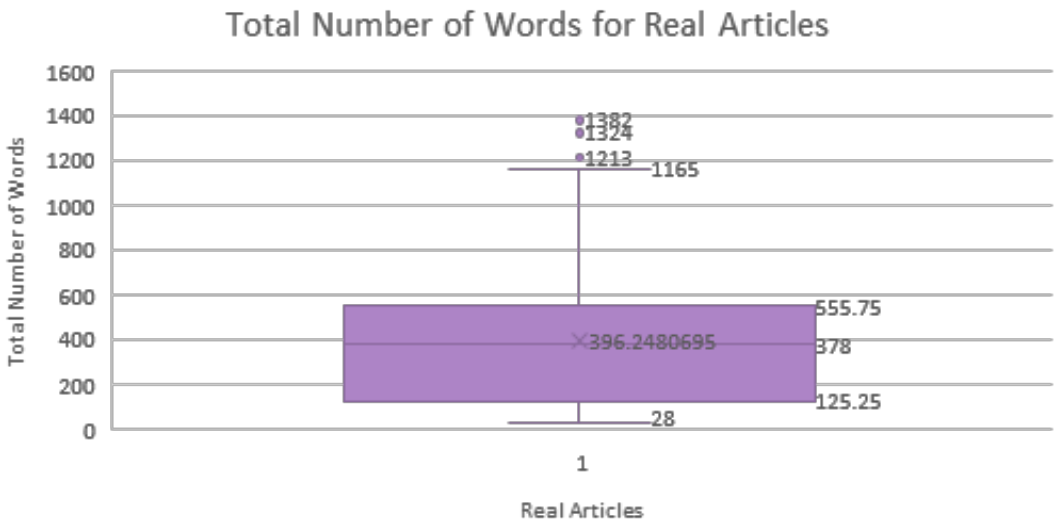
Are their indicators in the text of a news article that can be used by a machine learning algorithm to tell if it is real or fake article?

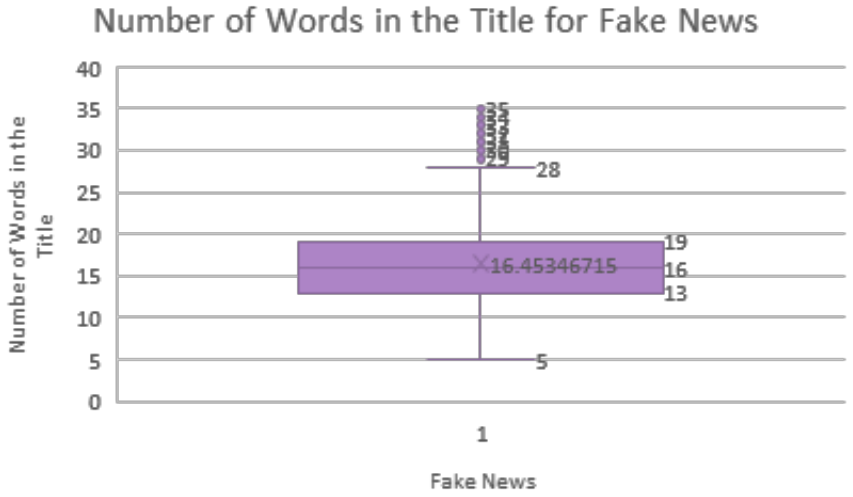
Yes, there are differences in the syntax and grammar of real news articles and fake news articles that can be used by a machine learning algortithm.

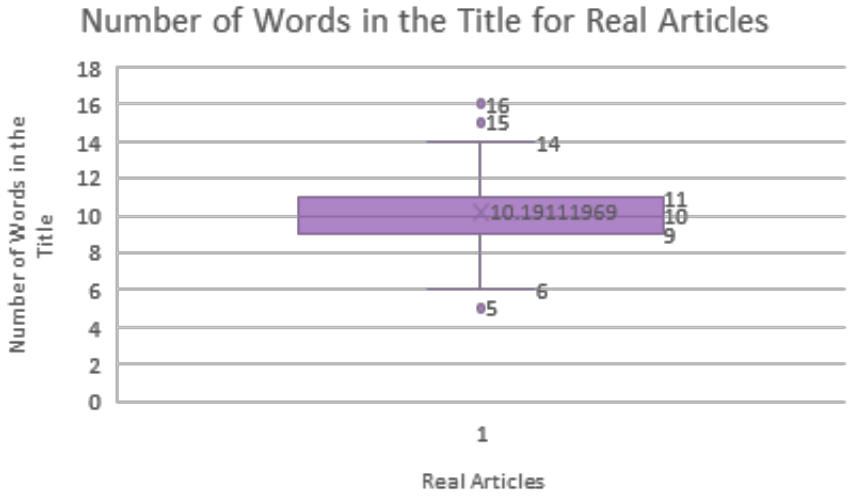
Additional Insights

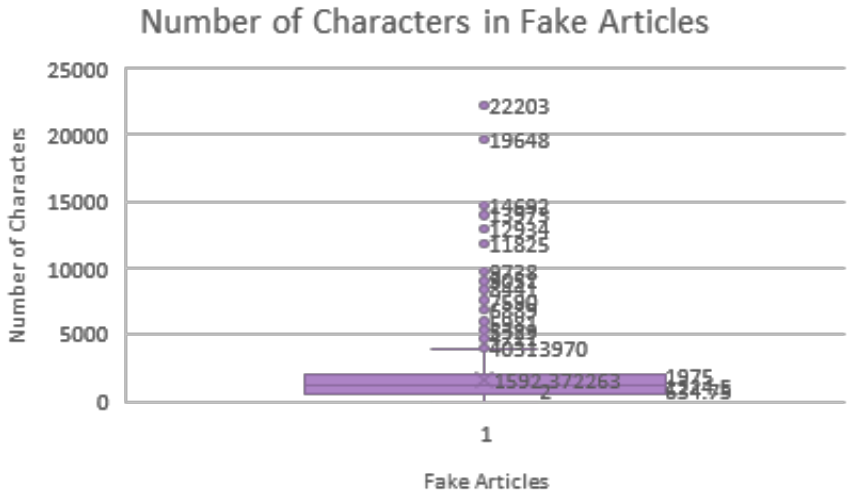
Real news articles have a consistent syntax format compared to fake news.

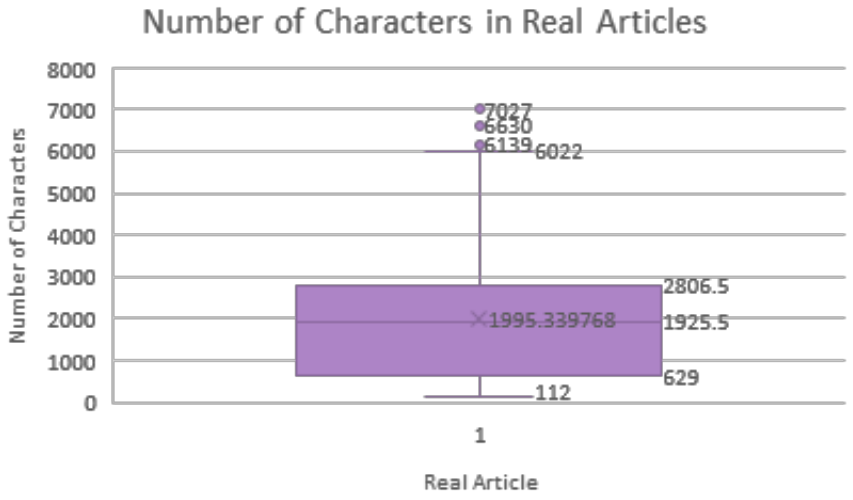





Number of Words in the Title for Fake News
Number of Words in the Title
40
35
30
25
20
15
10
5
0
28
19
16.45346715
16
13
5
1
Fake News


Number of Words in the Title for Real Articles
Number of Words in the Title
18
16
14
12
10
8
6
4
2
0
16
15
14
10.19111969
11
10
9
6
5
1
Real Articles


Number of Characters in Fake Articles
Number of Characters
25000
20000
15000
10000
5000
0
22203
19648
14692
11825
40313970
1592.372263
1975
1
Fake Articles


Number of Characters in Real Articles
Number of Characters
8000
7000
6000
5000
4000
3000
2000
1000
0
7027
6630
6139
6022
1995.339768
2806.5
1925.5
629
112
1
Real Article

Social Engineering: *The Art of Human Hacking*

Presented by: Devanshi Patel

RESEARCH QUESTION

What techniques do social engineers employ to influence their targets to carry out a successful social engineering attack?

MOTIVATION

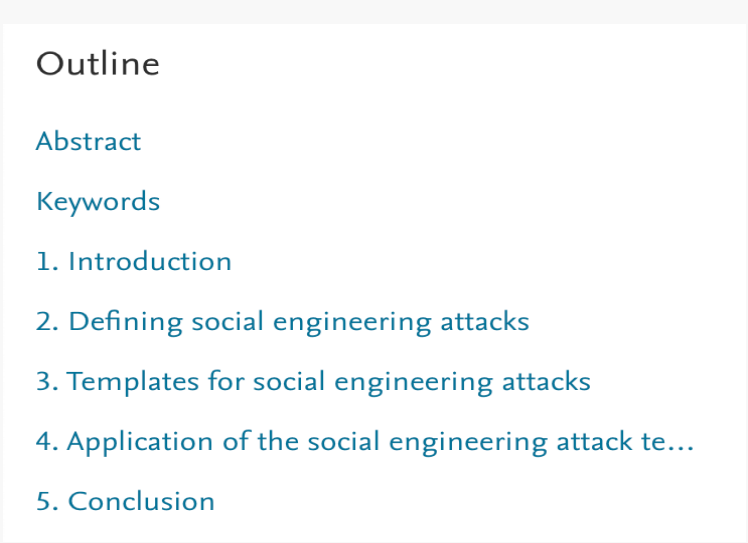
- Elevating Threat
- Effective countermeasures
- Raising Consciousness

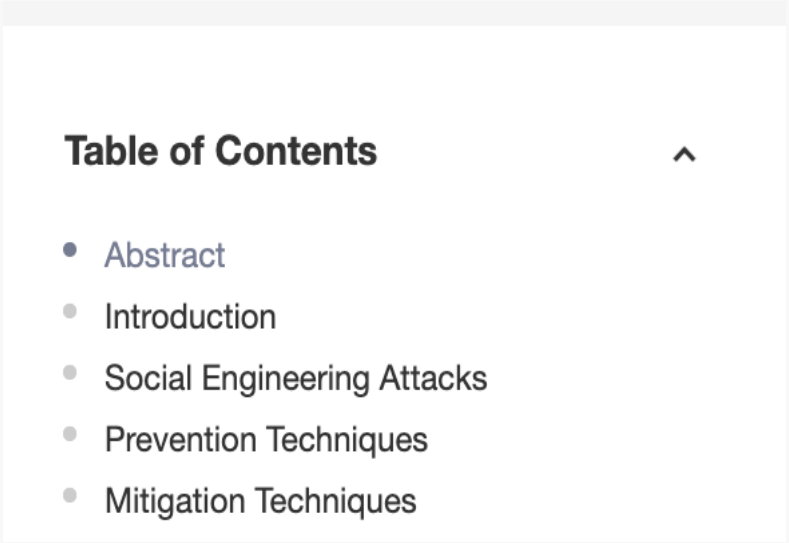
IMPLICATIONS

- Effective Awareness Training Programs
- Improved Organization Guidelines.

THINGS WE ALREADY KNOW: PREVIOUS STUDIES

- What it is?
- Common Types of Attacks
- Typical Attack Cycle
- Prevention Strategies





MY CONTRIBUTION

- Profound Attack Vector & Framework
- Psychological Principles Utilized
- Types of Emotions Exploited
- Target selection & Factors considered
- Technological Advancements/Future Trends

Small Businesses vs. Big Business

Sana Shaukat

Research Question, Motivation & Implications

- **RQ: What differences are there in cyberattacks between small and large businesses?**
- **Motivation:** Why is this an important question to study?
  - small businesses tend to have more cyberattacks when compared to bigger businesses
  - this capstone will also show how much more significant an attack is when it is on a smaller company
- **Implications:** Describe things you could do once you know the answer
  - Brings awareness to how these attacks impacts businesses differently
  - identifying which ones will help businesses protect themselves better

Things we know: Previous studies

- Describe what we already know about the RQ

  (include references to the academic papers / reports / investigations)
  - In a small business jobs such as information security are not always a priority, and can often be a part time job of a single person (Bedwell, 2014)
  - 20% of small companies rely on their security business unit vs. 62% of larger organizations (Pricewaterhouse Coopers, 2014)
  - " Fifty percent of small to medium sized businesses have been the victims of cyber attacks and over 60% of those attacked go out of business" – Dr. Jane Leclair, Chief Operating Officer at the National Cybersecurity Institute

- Your Contribution
  - There are a variety of studies on small business, but very few comparing big and small businesses
  - Not many papers comparing total revenues to cost of breaches
  - This capstone is a research design that focuses on the differences of business types by performing a t test

Plan VS Implementation

Adjusted Plan

- ❑ **Task 1: Most common cyberattacks on small and big business**
- ❑ **Task 2: Total revenue and costs of data breaches**
- ❑ **Task 3: Performing the T-test**

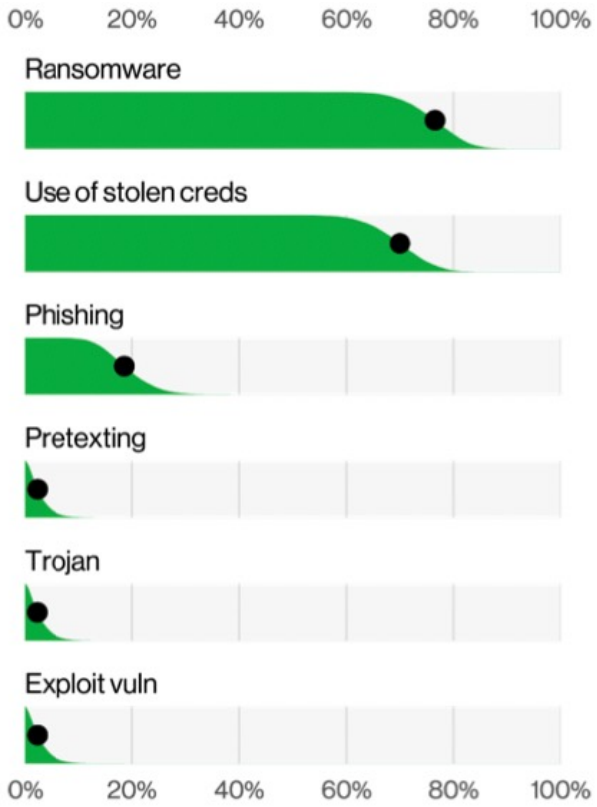
- ❑ **Initial RQ: Which Cyberattacks do small businesses face the most?**

Task 1

- Step 1
  - I first focused on the smaller business aspect, and what attacks were most common
  - Verizon's Data Breach Report from 2022 was my main point of collection for small businesses
- Step 2
  - Next, I did the same thing but for bigger businesses, and found data on Ponemans institute global risk report
  - I collected the top 6 for both small and big businesses, because of how extensive Ponemans list was

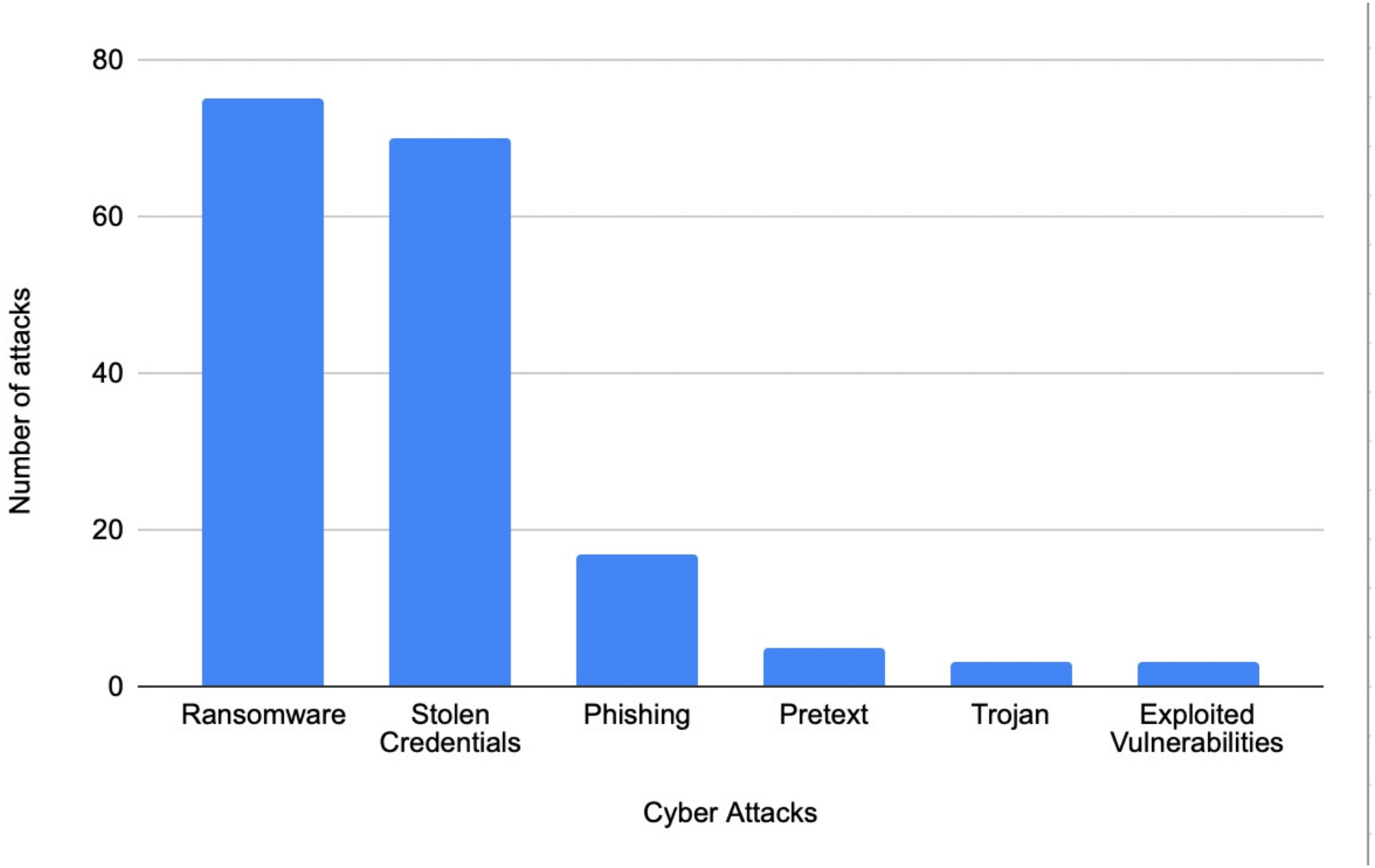
Results on next slides

Data Collection

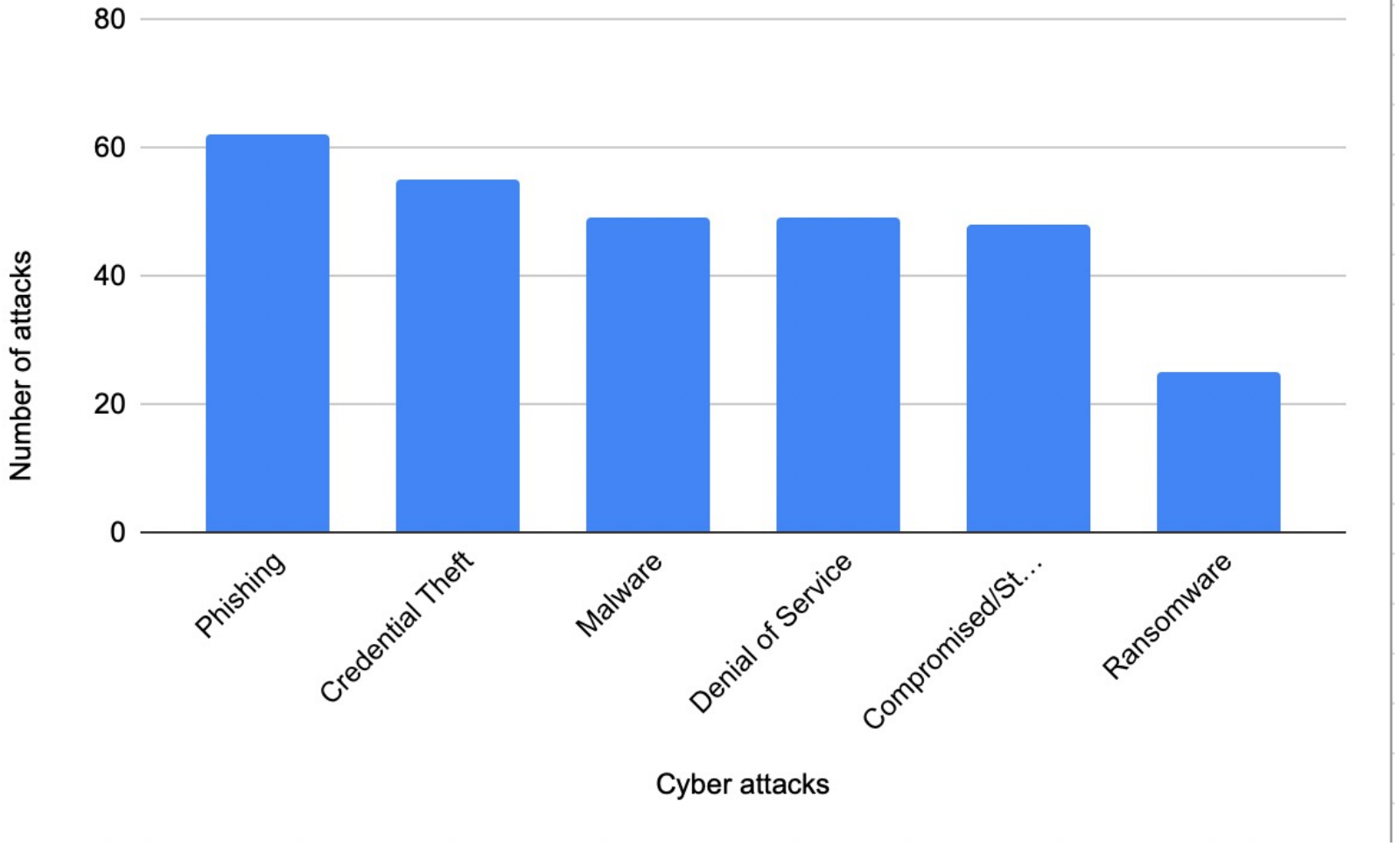


Account takeover	49%
Advanced malware / zero day attacks	29%
Compromised / stolen devices	48%
Credential theft	55%
Cross-site scripting	27%
Denial of service	49%
General malware	49%
Malicious insider	45%
Phishing / social engineering	62%
Ransomware	25%
SQL injection	21%
Web-based attack	28%
Other (please specify)	6%

Small Business Most Common Cyberattacks



Big Business Most Common Cyberattacks



Task 2

- Step 1
  - Used Comparitech data to find ransomware business cases
  - Would find the total revenues of companies and classify them as small or large business
- Step 2
  - After gathering several ransomware cases, I also focused on phishing attacks and the costs of those breaches
  - Inputting those results into excel and making charts for each

Data Collection

Year	Month, Year	Company Affected	Industry	Sub-Industry	City/County	State	# Records Affected	Ransom Paid	Ransom Amount	Ransomware Strain
2022	Mar, 2022	Aetna Bridge Company (ABC)	Business	Construction	Warwick	Rhode Island	108	Unknown	Unknown	LockBit
2021	May, 2021	All American Asphalt	Business	Construction	Corona	California	Unknown	Unknown	Unknown	Darkside
2021	Mar, 2021	Anvil Corporation	Business	Construction	Bellingham	Washington	667	Unknown	Unknown	Unknown
2021	May, 2021	Betenbough Homes	Business	Construction	Lubbock	Texas	Unknown	Unknown	Unknown	REvil
2022	Feb, 2022	BMS CAT, Inc.	Business	Construction	Haltom City	Texas	2,271	Yes	Unknown	Conti
2020	Sep, 2020	Building Material Distributors	Business	Construction	Galt	California	3,970	Unknown	Unknown	Unknown
2020	Sep, 2020	CB Masonry	Business	Construction	Little Rock	Arkansas	Unknown	Unknown	Unknown	Conti
2022	Aug, 2022	CentiMark	Business	Construction	Canonsburg	Pennsylvania	3,805	Unknown	Unknown	Unknown
2020	Jan, 2020	Civil & Building North America	Business	Construction	Miami	Florida	Unknown	Unknown	Unknown	Unknown
2021	Jul, 2021	Coghlin Electrical Corp.	Business	Construction	Worcester	Massachusetts	2,040	No	Unknown	Unknown

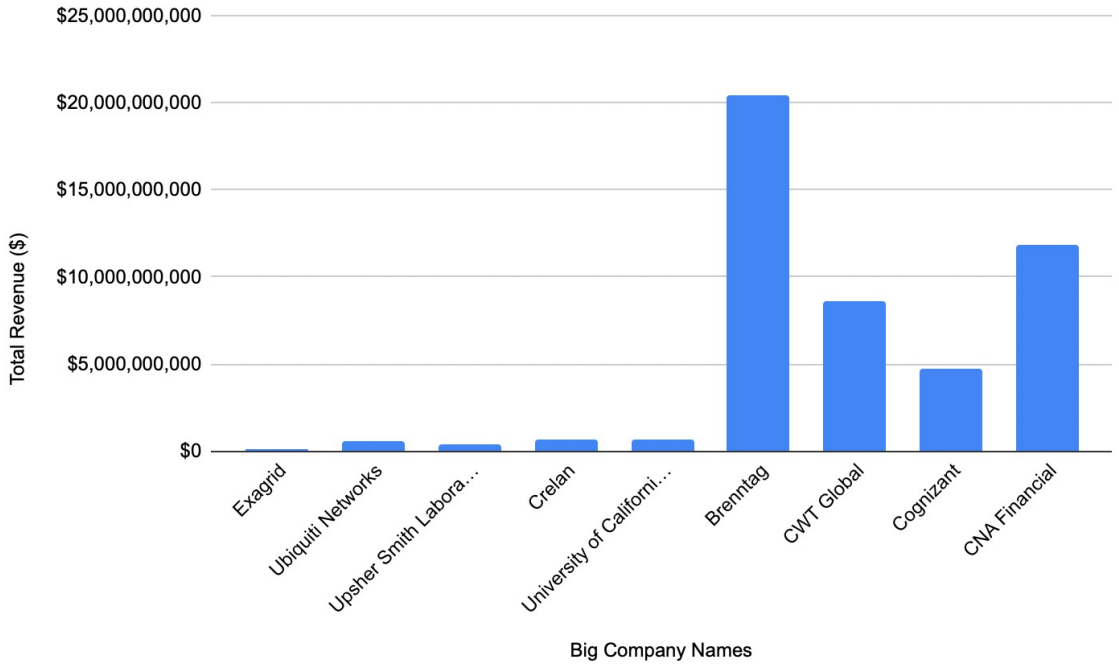
Data Collection

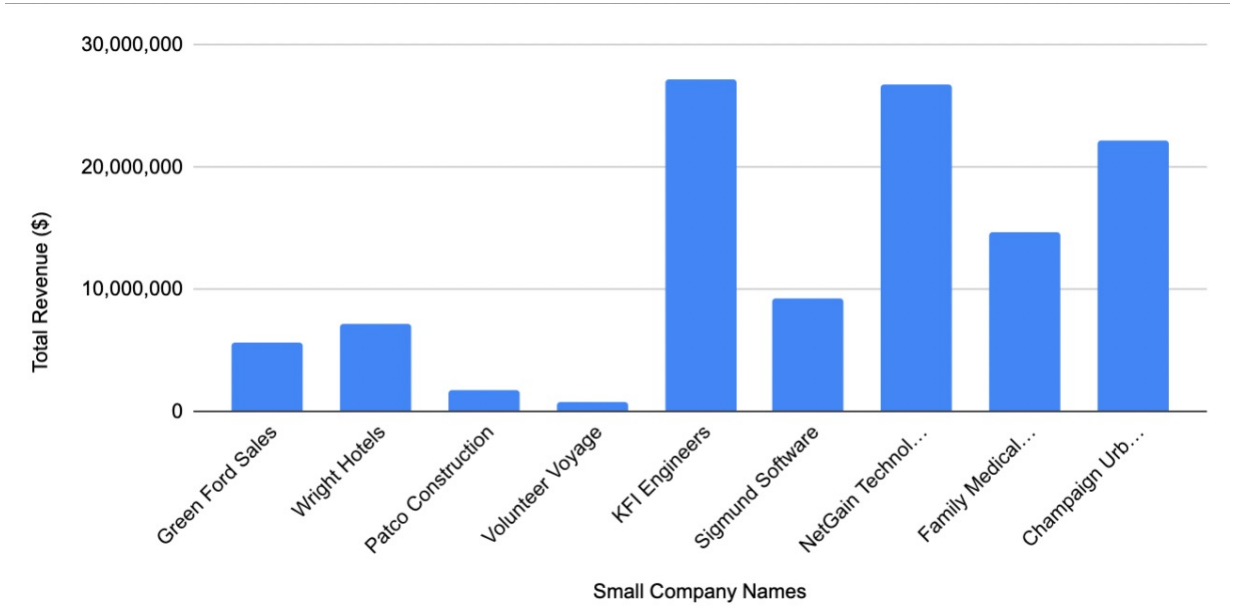
Company Name	Total Revenue	Cost of Data Breach
Green Ford Sales	5,600,000	$23,000
Wright Hotels	$7,200,000	$1,000,000
Patco Construction	$1,800,000	$588,000
Volunteer Voyage	$791,000	$14,000
KFI Engineers	$27,100,000	$300,000
Sigmund Software	$9,200,000	675,000
NetGain Technologies	$26,700,000	$2,300,000
Family Medical Center of Michigan	$14,717,459	$30,000
Champaign Urbana Public Health District	$22,100,000	$300,000

Data Collection

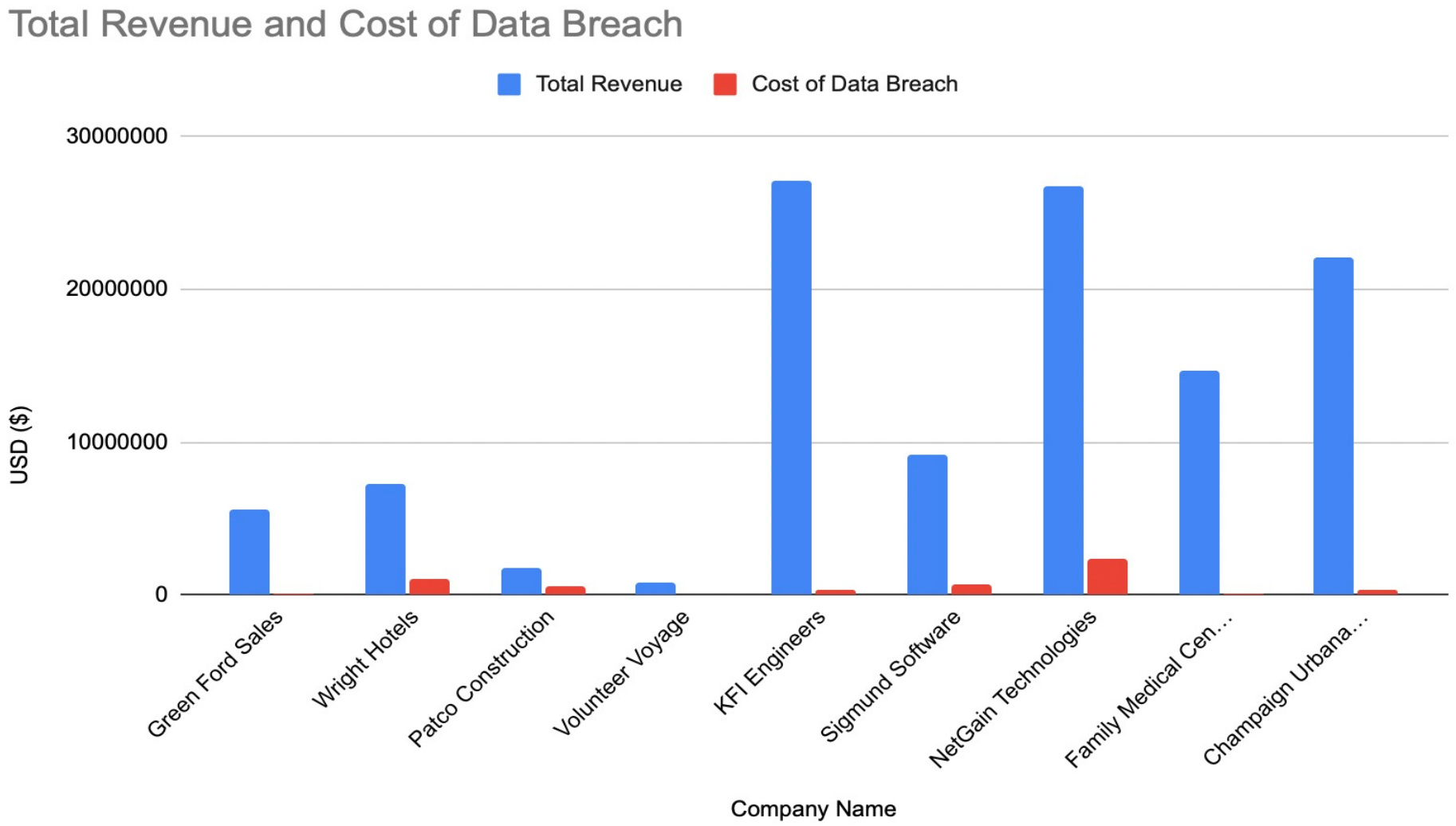
Company Name	Total Revenue	Cost of Attack
Exagrid	$69,200,000	$2,600,000
Ubiquiti Networks	$596,000,000	$39,000,000
Upsher Smith Laboratories	$420,000,000	$50,000,000
Crelan	$670,292,821	$76,791,400
University of California San Francisco	$640,000,000	$1,140,000
Brenntag	$20,430,000,000	$4,400,000
CWT Global	$8,600,000,000	$4,500,000
Cognizant	$4,700,000,000	$60,000,000
CNA Financial	$11,879,000,000	$40,000,000

Total Revenues

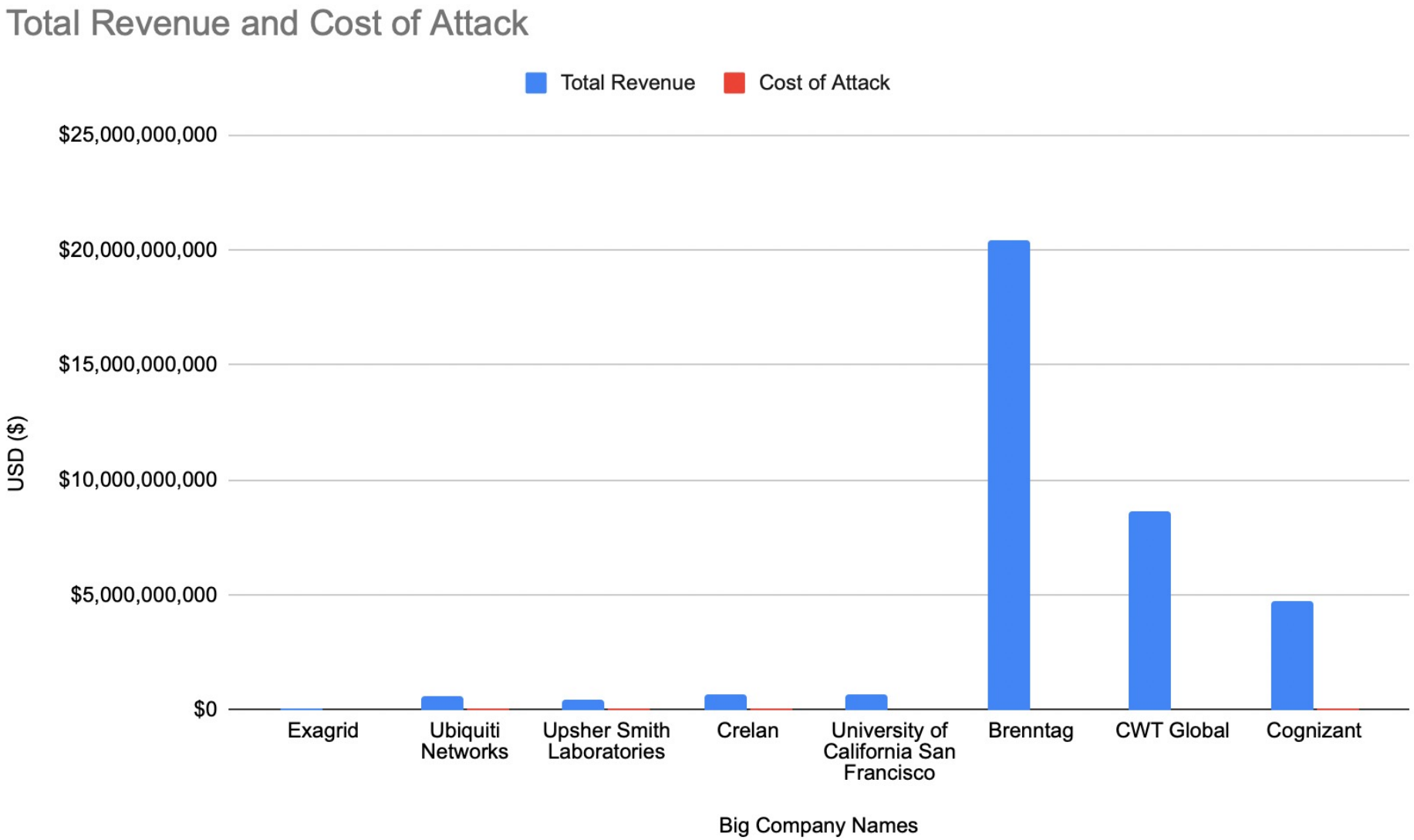




Small Business



Big Business



Task 3

- Step 1
  - With the data analysis tool in excel, created a T-test
  - Focused on the total revenue first to see if there was a statistical difference between those two factors
  - Then performed another T-test on cost of data breaches that were collected before

T-test on Total Revenue

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	*Variable 2*
Mean	12800939.89	5333832536
Variance	1.06155E+14	4.99866E+19
Observations	9	9
Pooled Variance	2.49934E+19	
Hypothesized Mean Difference	0	
df	16	
t Stat	-2.257822369	
P(T<=t) one-tail	0.019141947	
t Critical one-tail	1.745883676	
P(T<=t) two-tail	0.038283893	
t Critical two-tail	2.119905299	

T-test on cost of data breach

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	581111.1111	30936822.22
Variance	5.29223E+11	8.18976E+14
Observations	9	9
Pooled Variance	4.09752E+14	
Hypothesized Mean Difference	0	
df	16	
t Stat	-3.181162611	
P(T<=t) one-tail	0.002901438	
t Critical one-tail	1.745883676	
P(T<=t) two-tail	0.005802876	
t Critical two-tail	2.119905299	

Conclusions

- RQ:
  - **What differences are there in cyberattacks between small and large businesses?**
  - Statistical Differences can be seen throughout the graphs I have made, and when we look at the t test results for total revenue and cost of breach
- Additional Insights:
  - A lot of stats on small businesses, but not enough specific details
  - Growing emphasis on helping small businesses that I did not expect
  - Surprised that credential theft was top two for both big and small businesses
- Factors that can invalidate your results
  - Factor 1: sample size being small because of limited data, especially with small businesses
  - Factor 2: Cyber attack data changes yearly, this would only apply for 2022
  - Didn’t get to have a lot of variety in types of attacks, when comparing the two

Lessons I Learnt

Describe challenges you faced and how you plan to deal with them in future projects

- ❑ Challenge 1: Difficult to find specific data on small companies
- ❑ Ways to fix this: Send out my own survey to smaller companies
- ❑ Challenge 2: Difficulty sifting through literature reviews because sometimes there were opposing facts
- ❑ Ways to fix this: Making sure that papers I am looking at are up to date, and the information applies to the US
- ❑Challenge 3: Difficulty finding information on variety of cyberattacks

PLAN VS IMPLEMENTATION

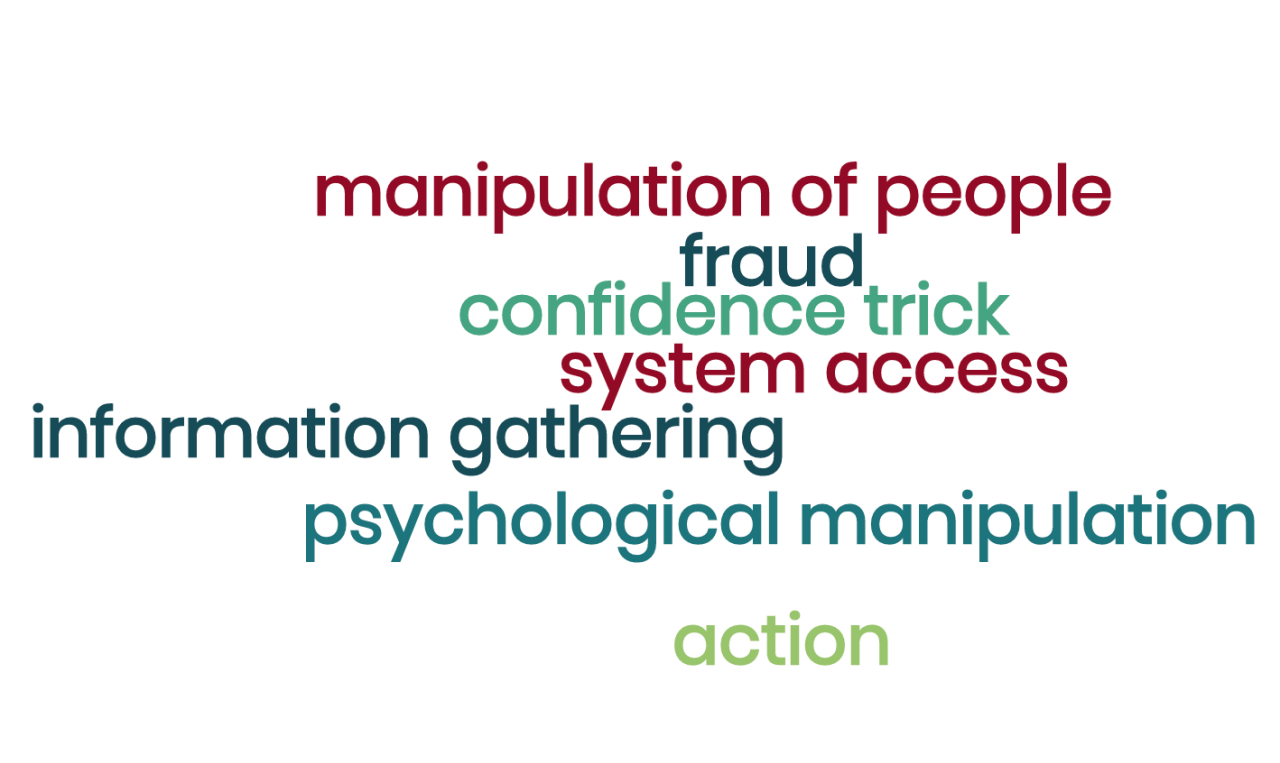
INITIAL PLAN

- Target selection & Factors considered
- Profound Attack Vector
- Psychological Principles Utilized
- Types of Emotions Exploited
- Technological Advancements/Future Trends
- **Social Engineering Attacks Dataset Analysis**
- **Survey Analysis on Awareness**

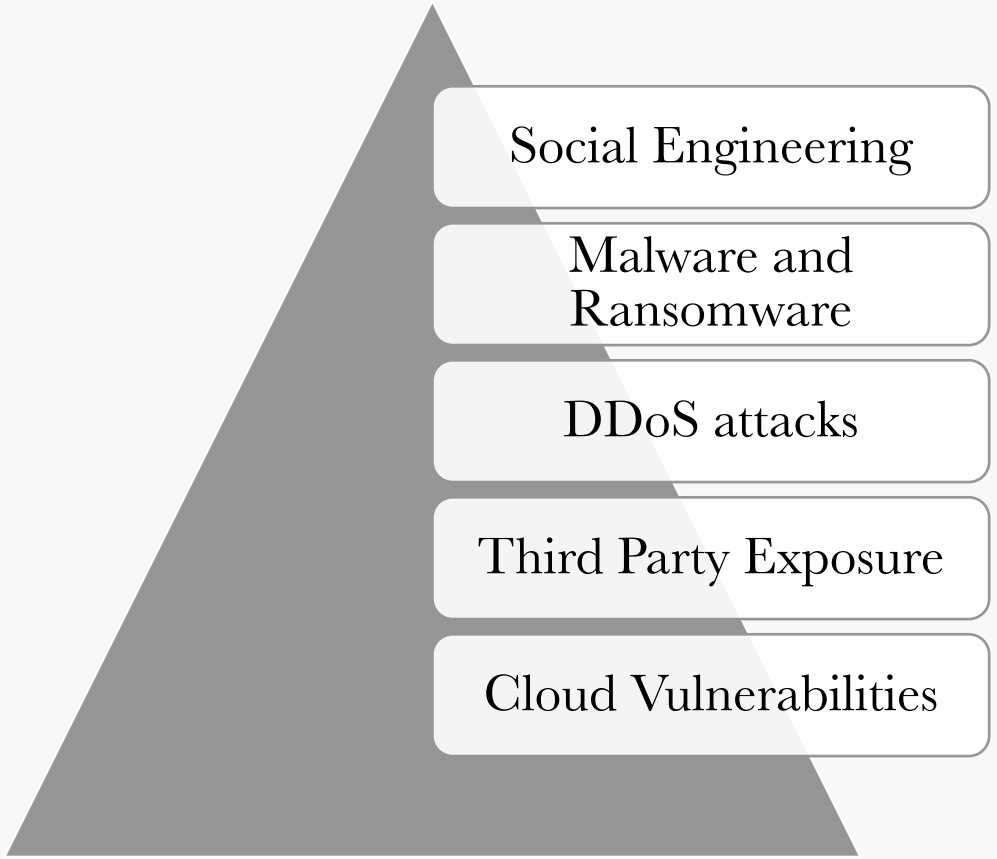
ADJUSTED PLAN

- Target selection & Factors considered
- Profound Attack Vector **& Attack Framework**
- Psychological Principles
- Types of Emotions Exploited
- Technological Advancements/Future Trends

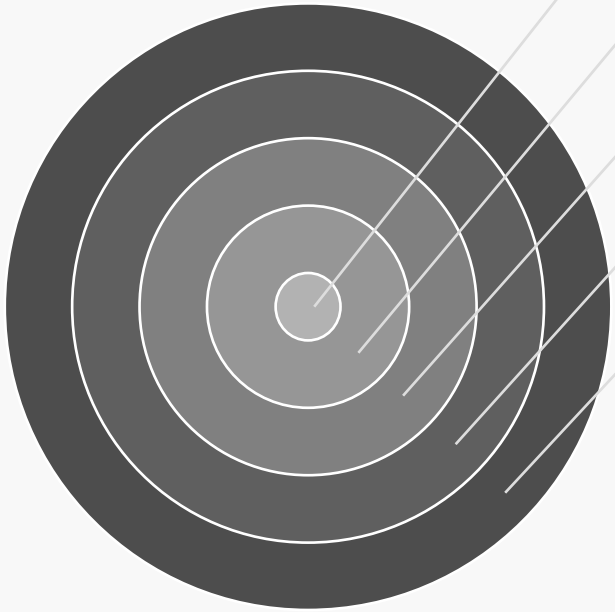
WHAT IS SOCIAL ENGINEERING



RANKING


Social Engineering
Malware and Ransomware
DDoS attacks
Third Party Exposure
Cloud Vulnerabilities

GOAL



Financial Gain

Identity & Data theft

Unauthorized access

Manipulate people

Service Disruption

CONSEQUENCES

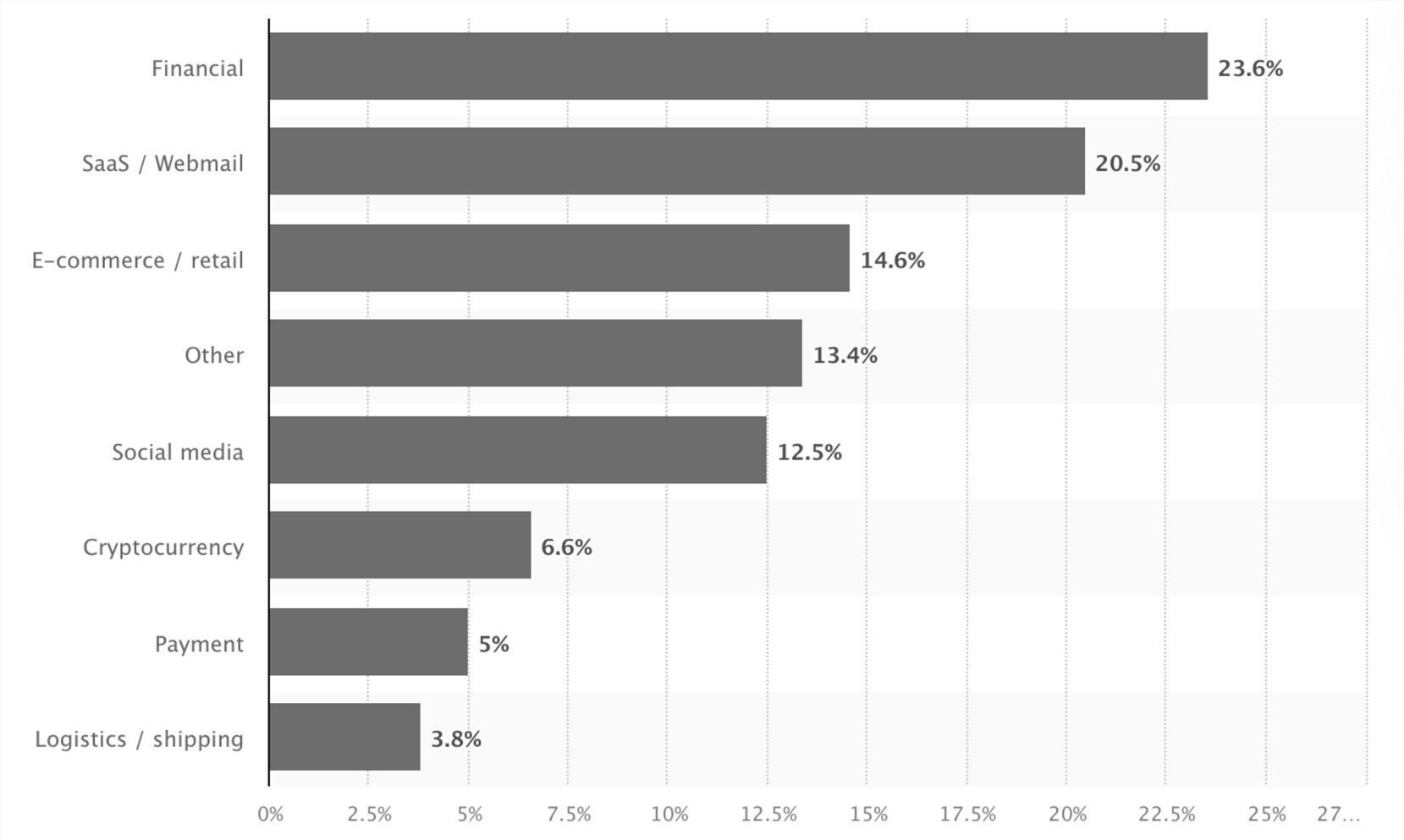
Malware attacks

Ransomware attacks

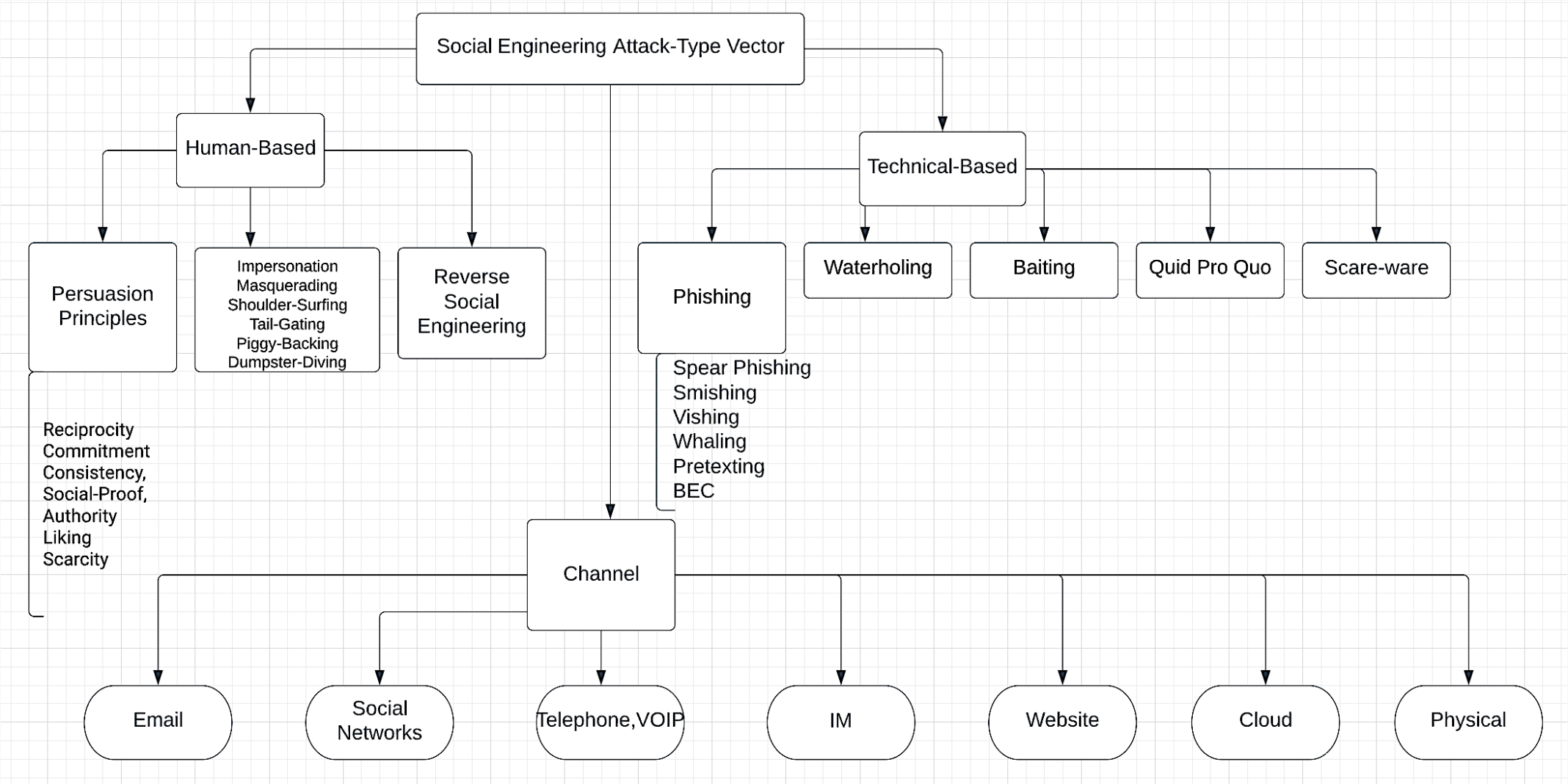
Reputational damage

Emotional Damage

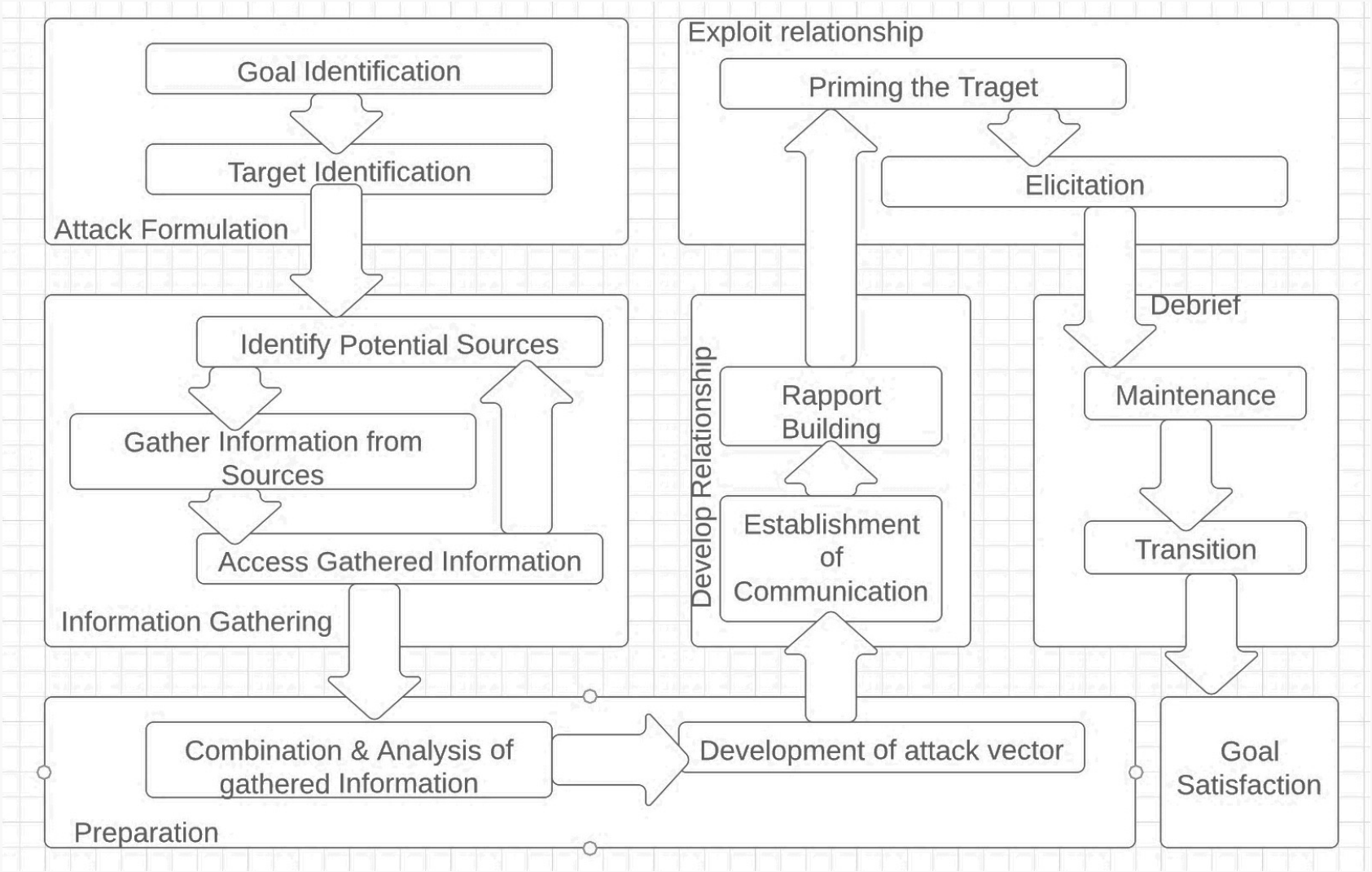
TARGET


Financial
23.6%
SaaS / Webmail
20.5%
E-commerce / retail
14.6%
Other
13.4%
Social media
12.5%
Cryptocurrency
6.6%
Payment
5%
Logistics / shipping
3.8%
0%
2.5%
5%
7.5%
10%
12.5%
15%
17.5%
20%
22.5%
25%
27...

ATTACK VECTOR/TAXONOMY

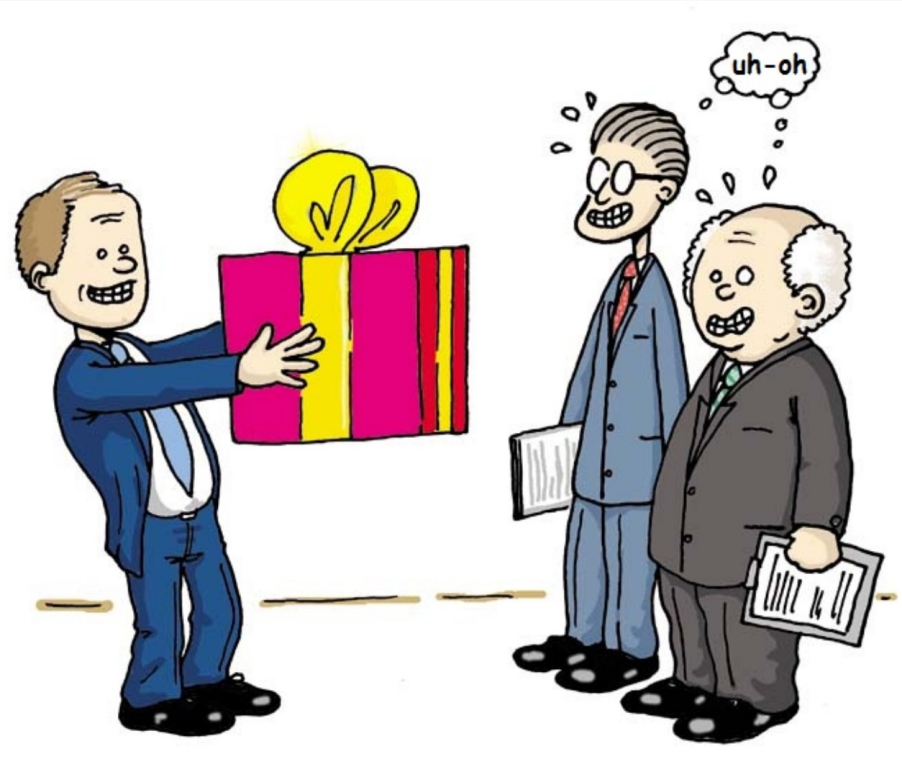


ATTACK FRAMEWORK



PSYCHOLOGICAL PRINCIPLES

RECIPROCITY



COMMITMENT AND CONSISTENCY

SOCIAL PROOF

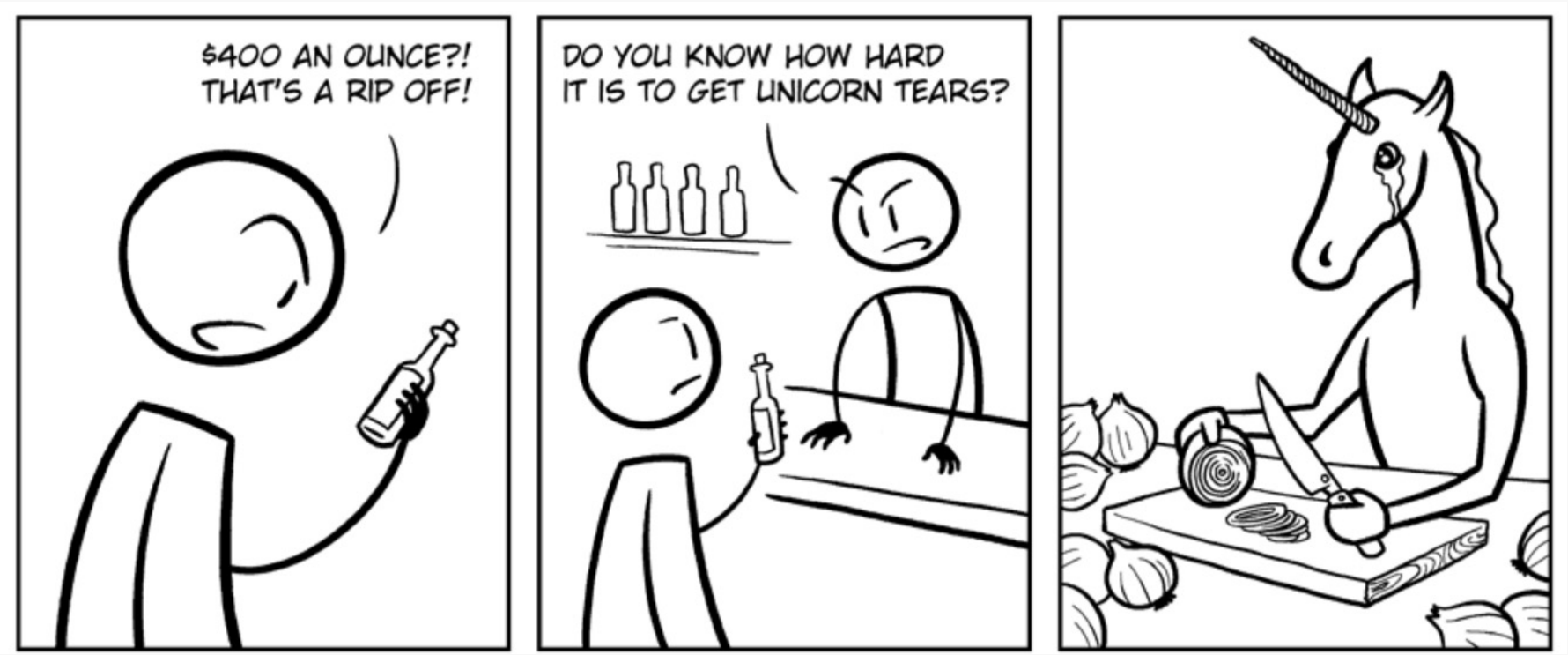
AUTHORITY


HARDIN
"I went into that meeting with a hidden agenda and then forgot where I put it."


NOBODY KNOWS ANYTHING.
WHAT A RELIEF.
THANK GOODNESS.
I THOUGHT IT WAS JUST ME.

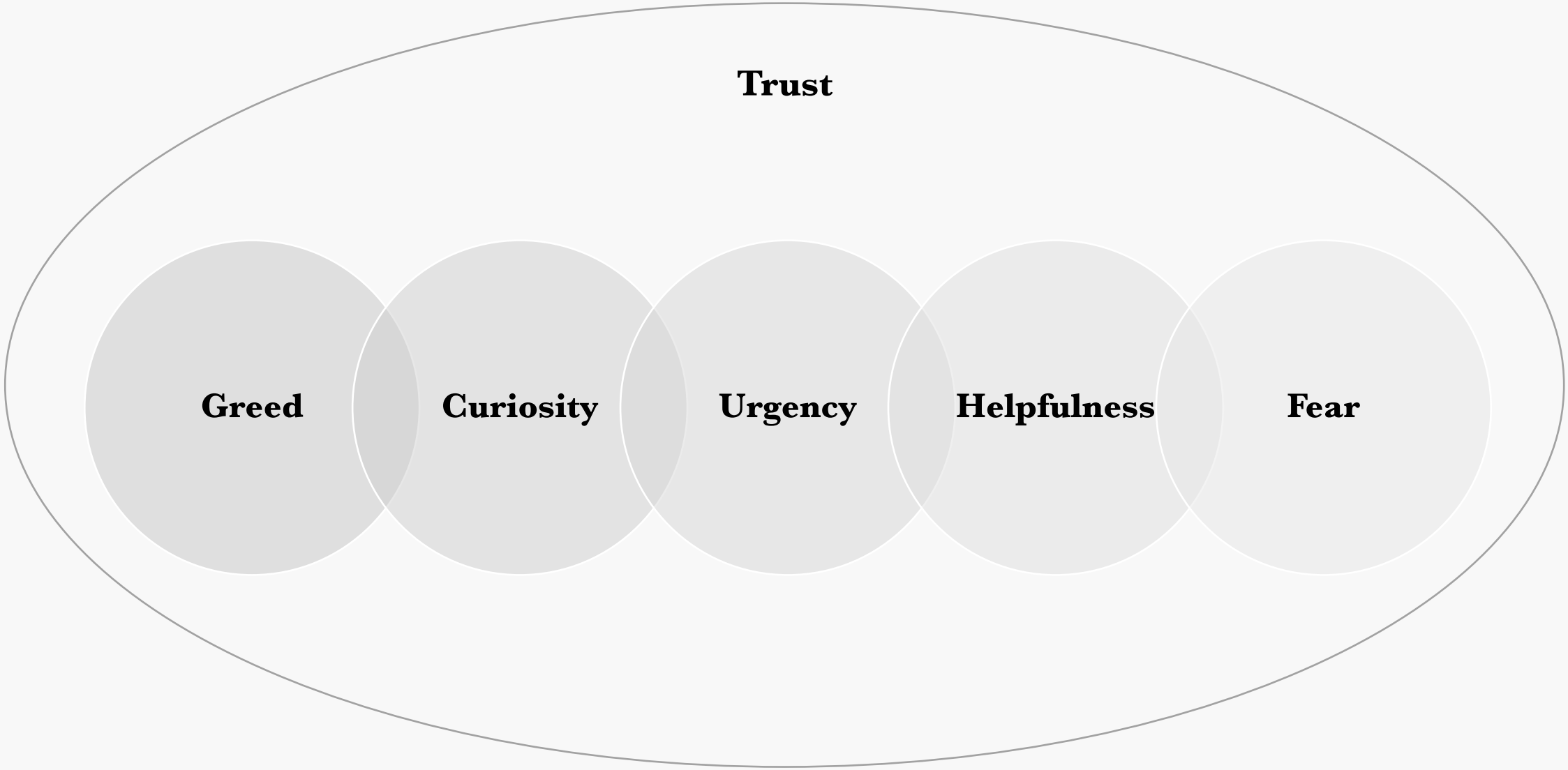
LIKING

SCARCITY


$400 AN OUNCE?!
THAT'S A RIP OFF!
DO YOU KNOW HOW HARD
IT IS TO GET UNICORN TEARS?


SALE

EMOTIONS EXPLOITED



GREED

Invitation to Google Analytics users

RR Rasmus Refer, Fastbase Inc. <newsletter@linkedupdates.com>
To **media**

Reply | Reply All | Forward | ...

Thu 5/16/2019 12:56 AM

If there are problems with how this message is displayed, click here to view it in a web browser.
Click here to download pictures. To help protect your privacy, Outlook prevented automatic download of some pictures in this message.

Web Version

Invitation

Good morning,

Last chance to reserve shares with discount in Fastbase Pre-IPO

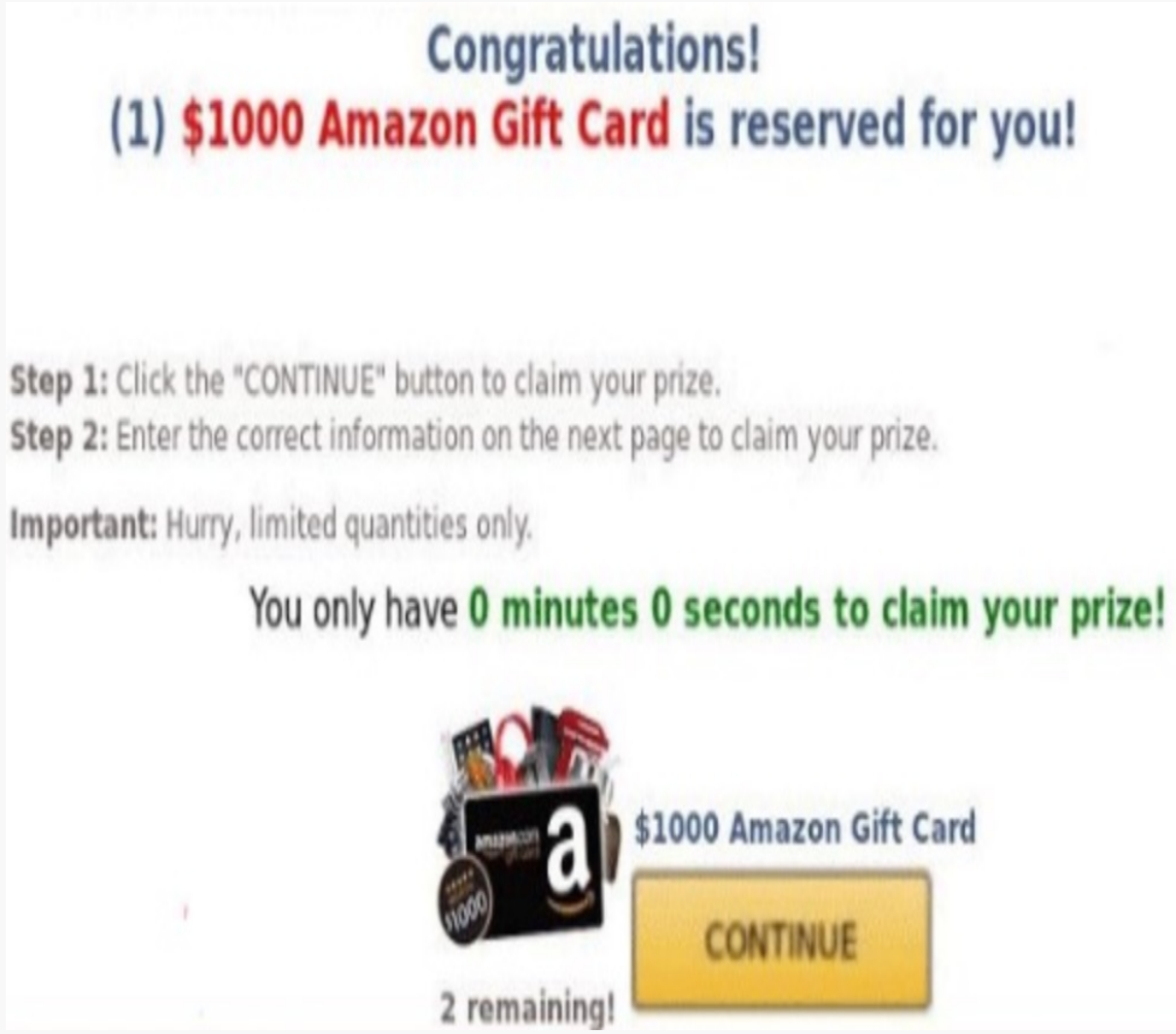
Since your company are using Google Analytics, you have the opportunity to purchase up to 25,000 shares at US$ 3.45 per share including discount.

Fastbase is targeting admission to the OTC Stock Market in June, 2019, as the fastest-growing SaaS web analytics and lead generation tool built upon Google Analytics.

Fastbase analyzes over 6 billion website visitors from over 1,000,000 companies and top brands around the world. Fastbase introductory 2019

The closing date for reserving shares is **May 15, 2019**.

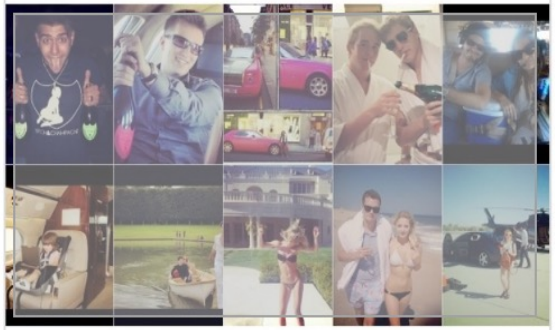
To reserve your shares now, use the share reservation form.
Share reservation



CURIOSITY

Wyatt OMG you're really on this Worst Instagram Wall. Someone put all your photos on there. Your currently ranked 10! this is so wrong. You can even see who added you on it.

Look



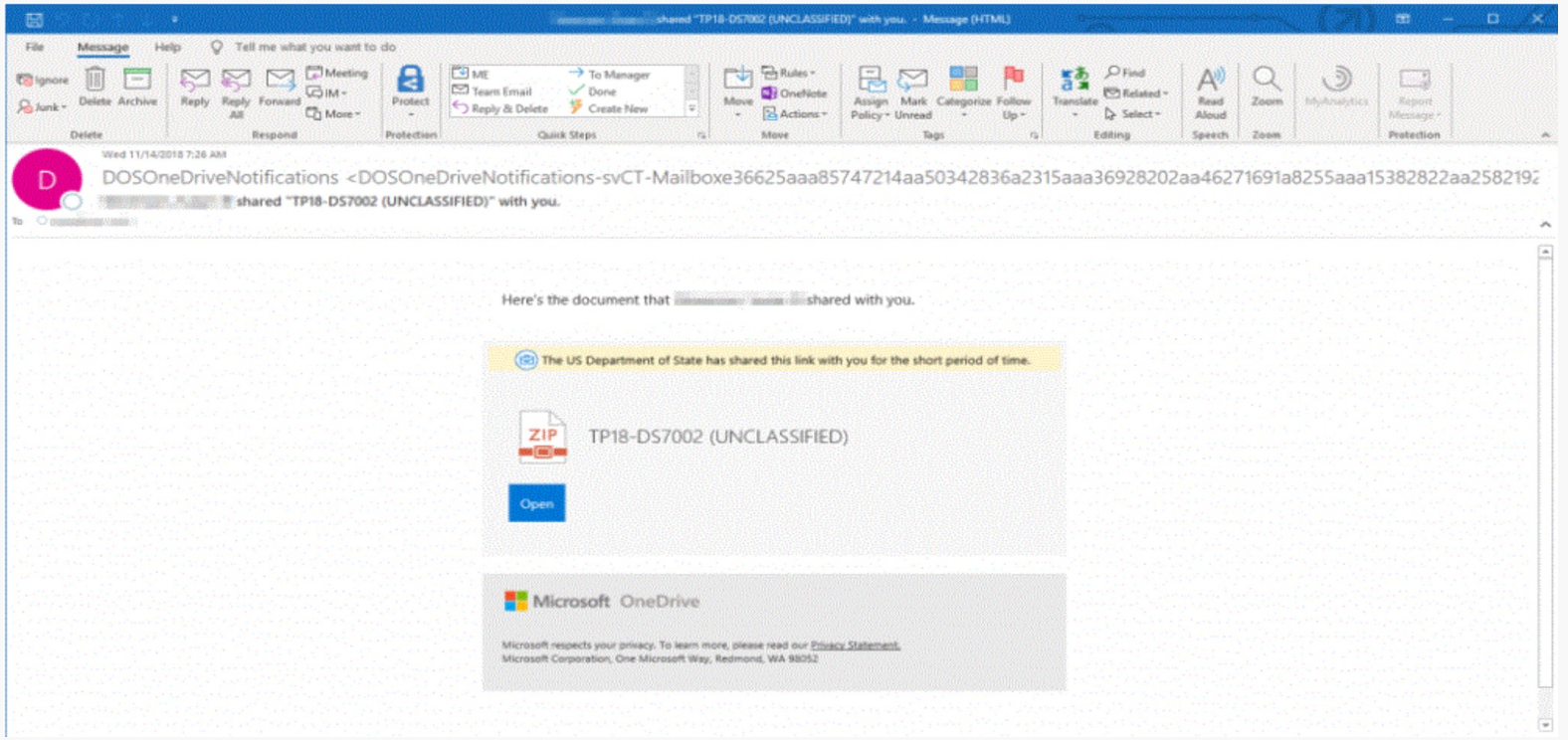
These Are the Worst People to Ever Use Instagram (Updated)

It takes a lot to find something new on the Internet that makes you hate our entire species, and yet, this: dozens of brats...

Message...

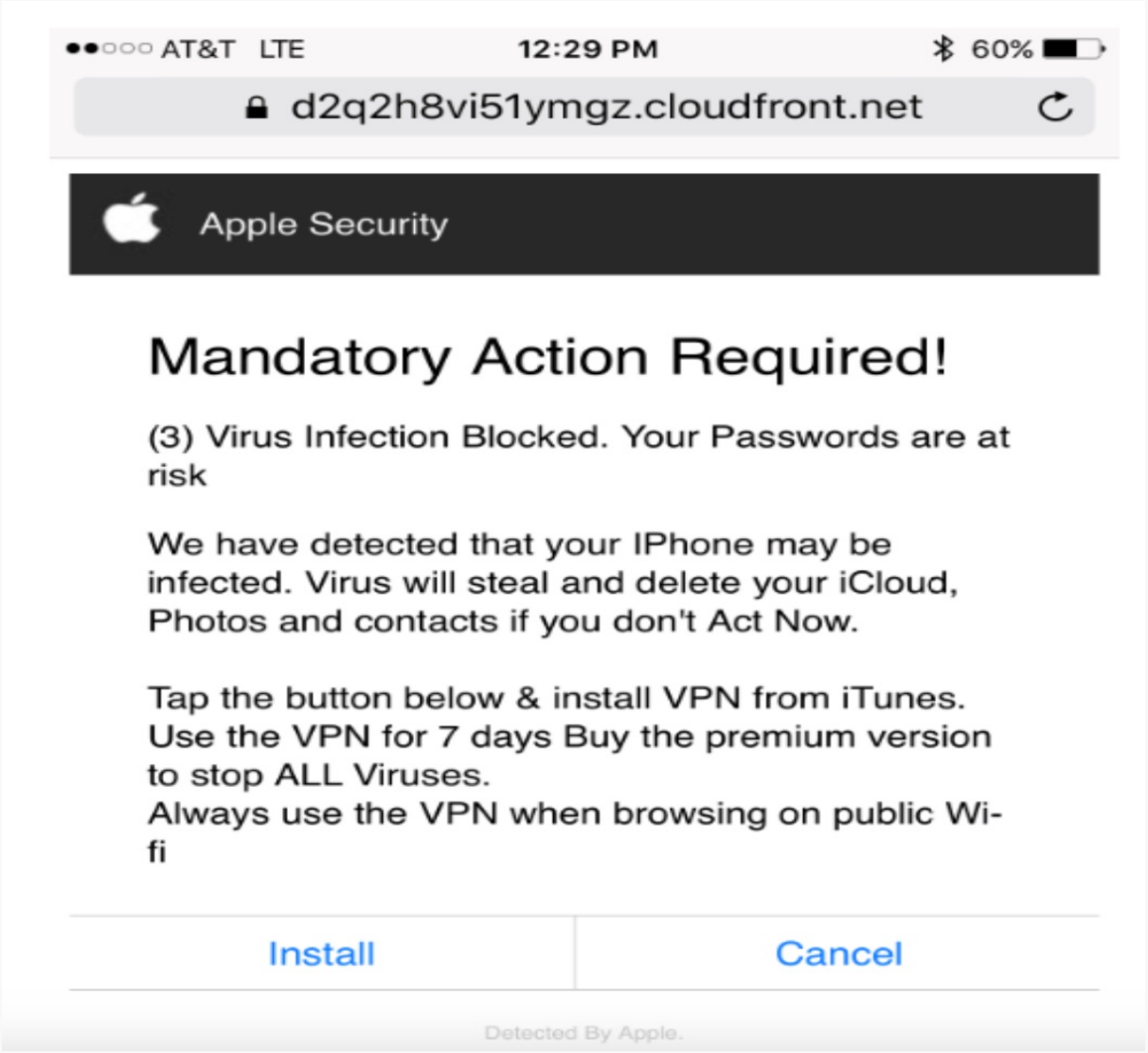
Text Message
Today 5:56 PM

Hello Olivia, your FEDEX package with tracking code GB-6412-GH83 is waiting for you to set delivery preferences: e3fmr.info/onAyXsVfomA



URGENCY


AT&T Wi-Fi
3:27 PM
72%
1 (410) 200-513
Text Message
Today 3:12 PM
FRM:AMEX@Message
SUBJ:Card Alert
MSG:Card Alert;Card Alert
http://
www.americanexpress-
message.com


AT&T LTE
12:29 PM
60%
d2q2h8vi51ymgz.cloudfront.net
Apple Security
Mandatory Action Required!
(3) Virus Infection Blocked. Your Passwords are at risk
We have detected that your IPhone may be infected. Virus will steal and delete your iCloud, Photos and contacts if you don't Act Now.
Tap the button below & install VPN from iTunes.
Use the VPN for 7 days Buy the premium version to stop ALL Viruses.
Always use the VPN when browsing on public Wi-fi
Install
Cancel
Detected By Apple.

HELPFULNESS

Ukraine Humaniterian Donation

From Ukraine / Україна Govt® on 2022-02-28 15:19

Details Plain text

A donation campaign has been launched to support Ukrain and also help refugees fleeing the conflict in Ukraine
The campaign, organized by the humanitarian organization Act for Peace, is hoping to raise **$9,000,000** to support refugees in the region.

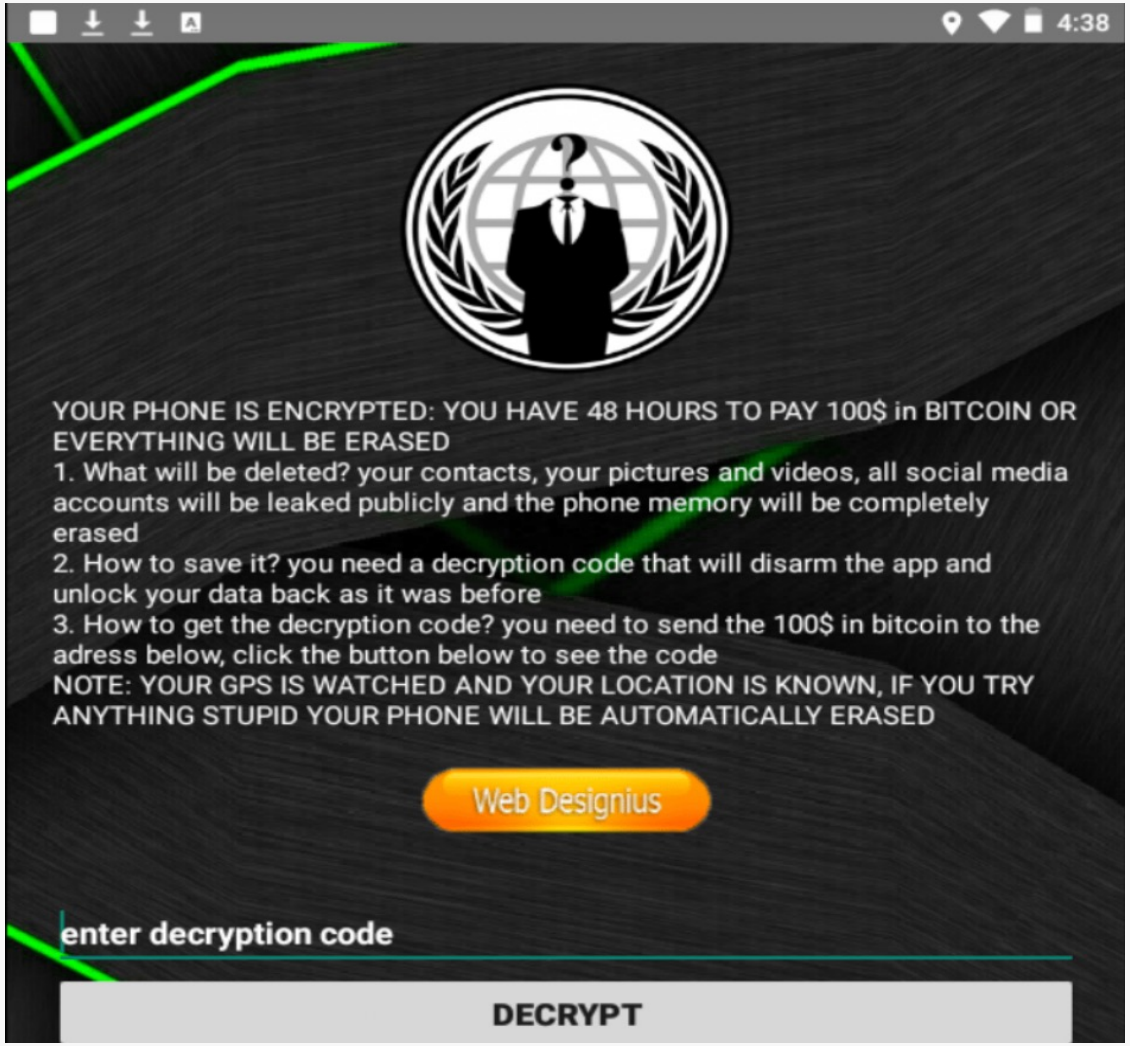
Stand with the people of Ukraine. Now accepting cryptocurrency donations. Bitcoin, Ethereum, USDT and NFT

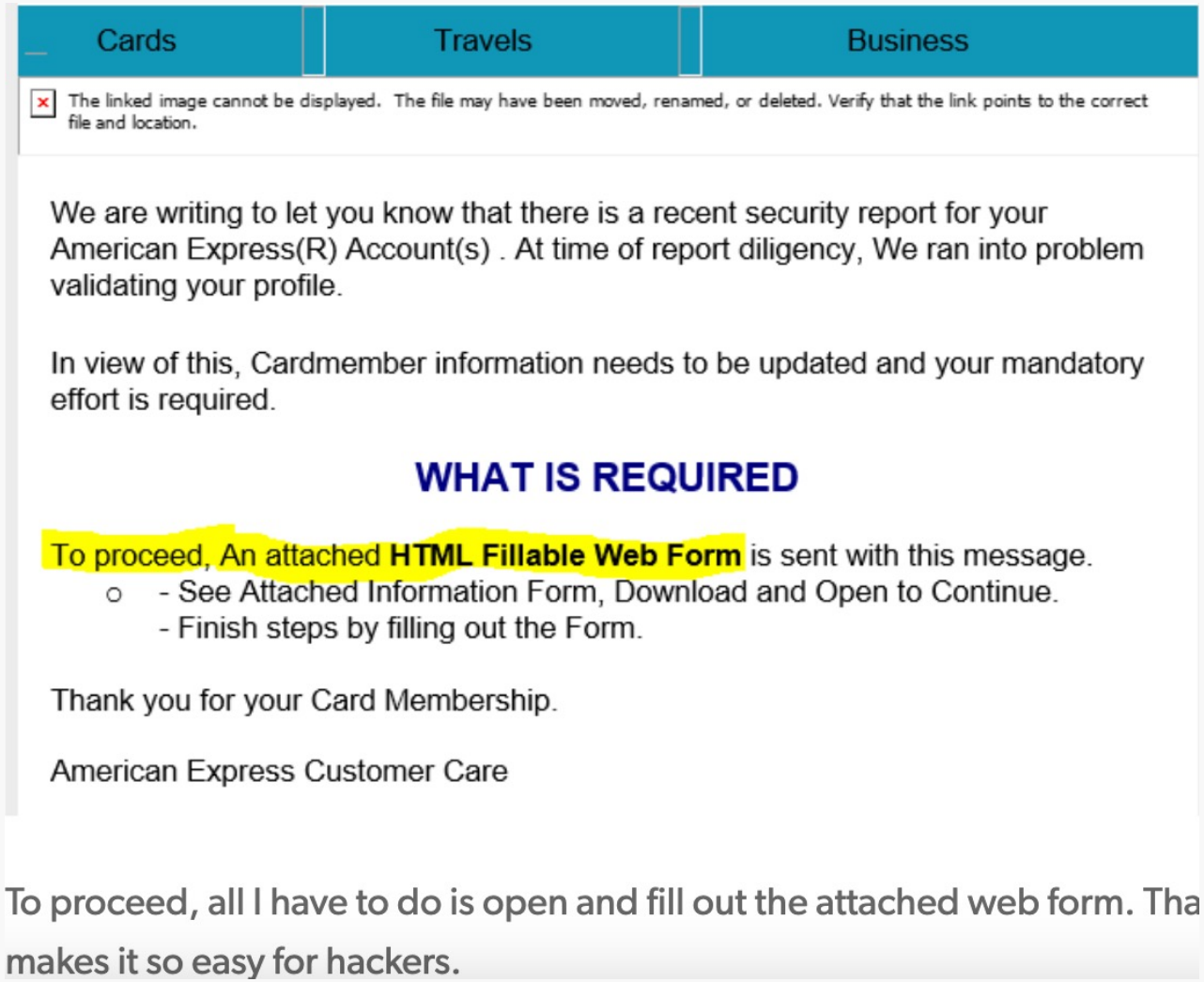
BTC- bc1qqglytupu8pup07eksh3gd77edkd3ge7ku6809y

ETH- 0x5535480a9D0F39b545F9c14b0a70a8755237b01f

Best Regards
Ukraine
#BeautifulUkraine

FEAR



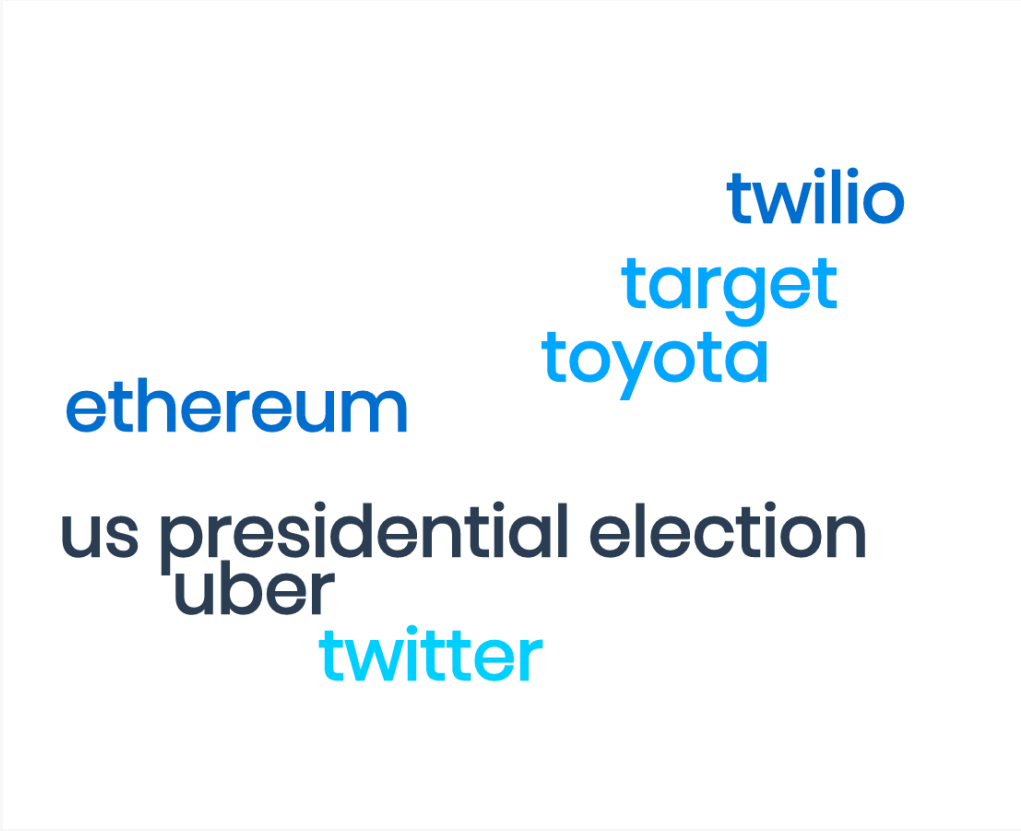


To proceed, all I have to do is open and fill out the attached web form. Tha makes it so easy for hackers.

TARGET SELECTION & FACTORS CONSIDERED

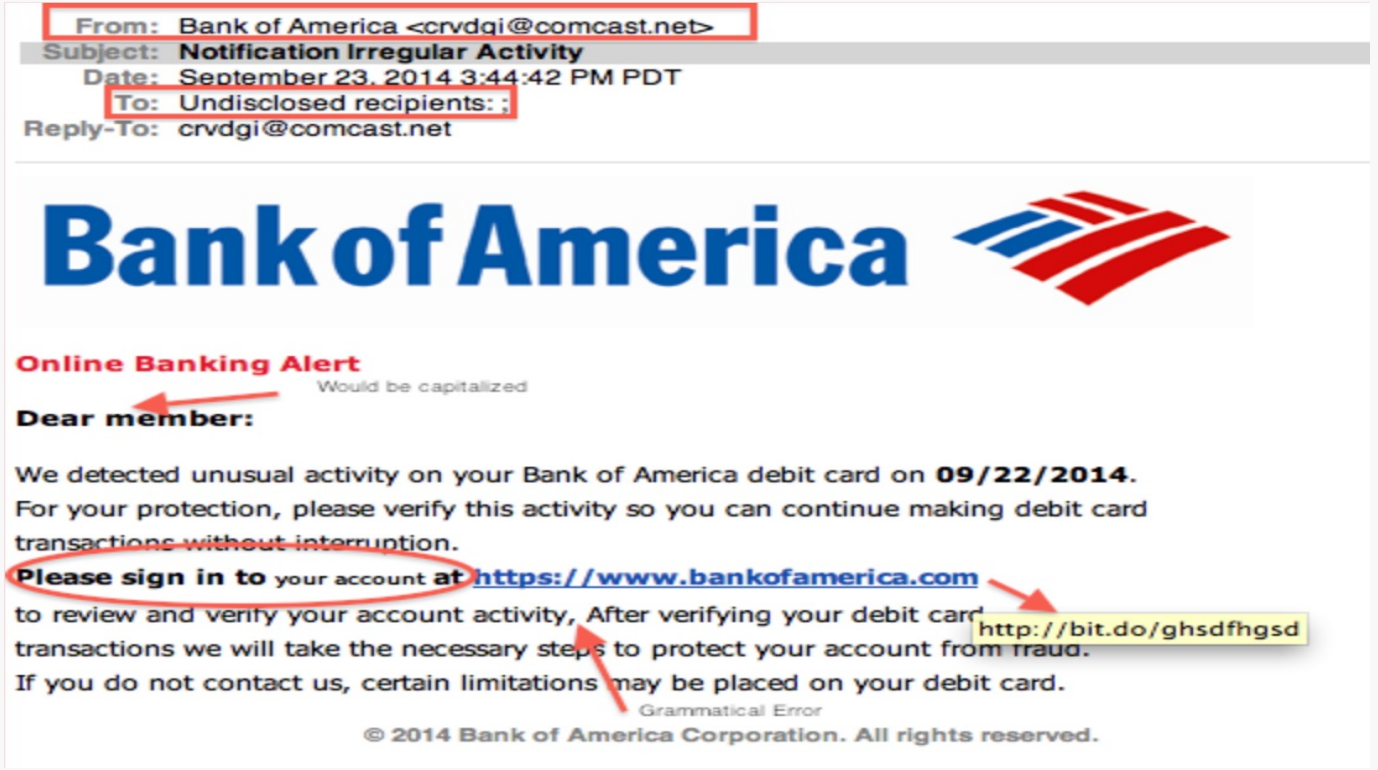
- Role and Access
- Job Title and Seniority
- Industry/Sector

CASE STUDIES



COUNTERMEASURES

- Check the source
- What do they know?
- Use a good spam filter
- Is this realistic?
- Don't go too fast



TECHNOLOGICAL ADVANCEMENTS/FUTURE TRENDS

WHAT'S COMING?

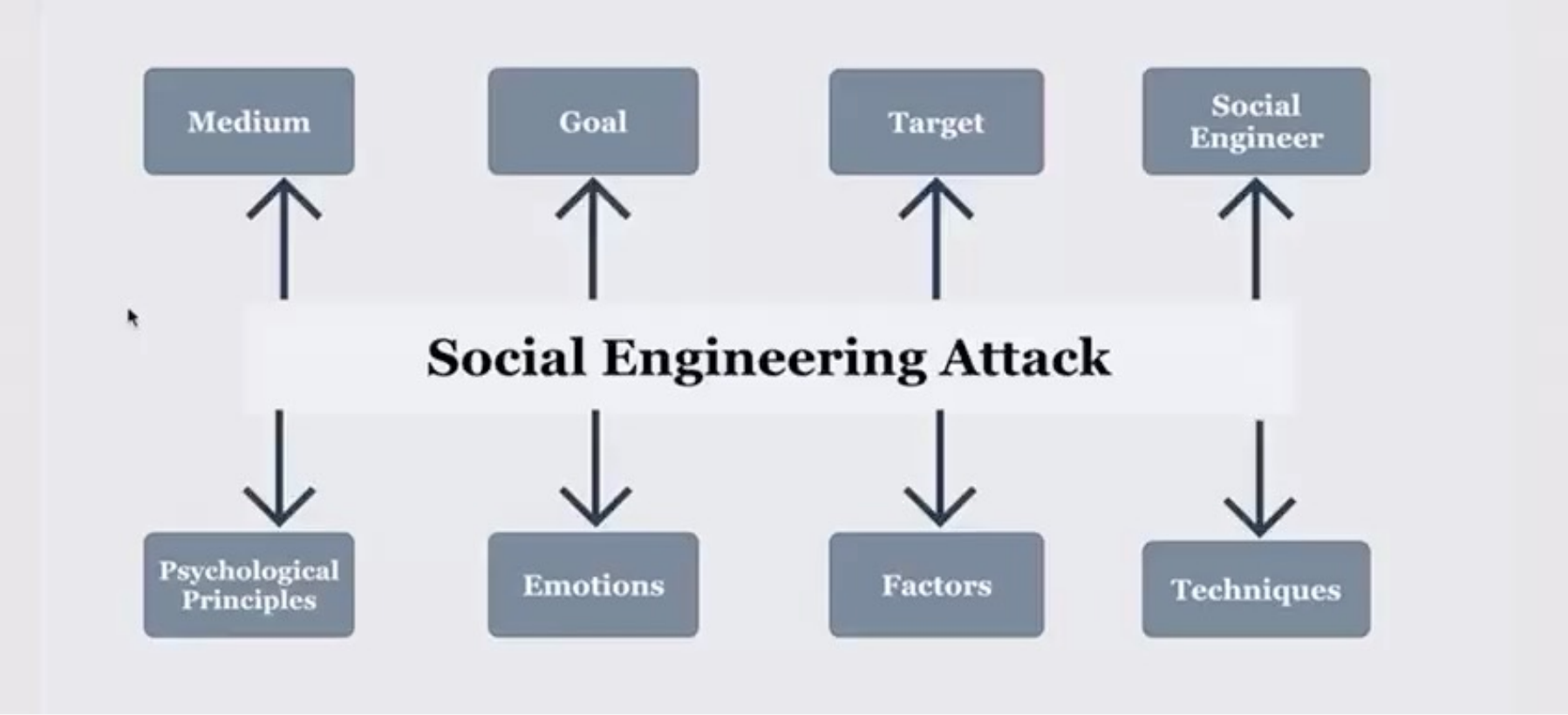
1. **Abuse of Law Enforcement Privileges**
2. **Deepfake Attacks**

FINAL THOUGHTS

LIMITATIONS/ CHALLENGES AND FUTURE PLAN

- Finding Attack dataset & Analyzing it.
- Survey Analysis on Awareness.

INTERACTIVE DIAGRAM



THANK YOU :)

How to Not Get Hacked

James Wrabel and Sayanta Pal Chowdhury

Research Question

How can we more effectively design ISA training to include cognitive psychology to achieve greater awareness?

Q1: Is ISA training proven to be effective?

Q2: Do current ISA trainings incorporate cognitive psychology?

Has incorporating cognitive psychology been shown to improve ISA? How can we measure this?

Q3

Motivation

- Human error is the most prevalent cyber threat
- Maybe we are asking the wrong questions – studies show individual differences play a factor in ISA
- Lack of research into current ISA training platforms and how they can be improved

Implications

- Proposal of a new form of ISA training that is more effectively designed to reduce human error

Q1 – Is ISA Training proven to be effective?

Yes… kind’ve

“Training can be effective… but”

- “there is a need for more rigorous evaluation”
- “effectiveness varied depending on the delivery method and content of the training”
- “effectiveness was limited by a lack of motivation and reinforcement”
- “effectiveness was limited by a lack of reinforcement and integration with other security controls”
- “additional studies should be conducted”

Ideally…

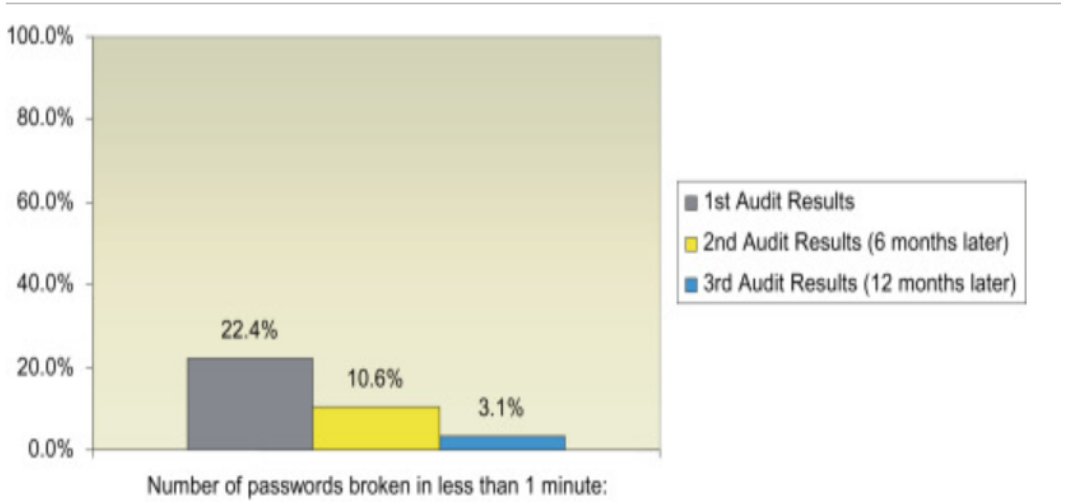
- An effective way to measure user knowledge, behavior, and judgement before and after ISA training
  - Testing user security habits
  - Simulating social engineering labs
  - Role playing exercises
  - Pen testing and incident response exercises
- Egelman and Peer (2015) stated that “there exists no standard measurement tool for end-user security behaviors”
- Need to avoid self-reporting

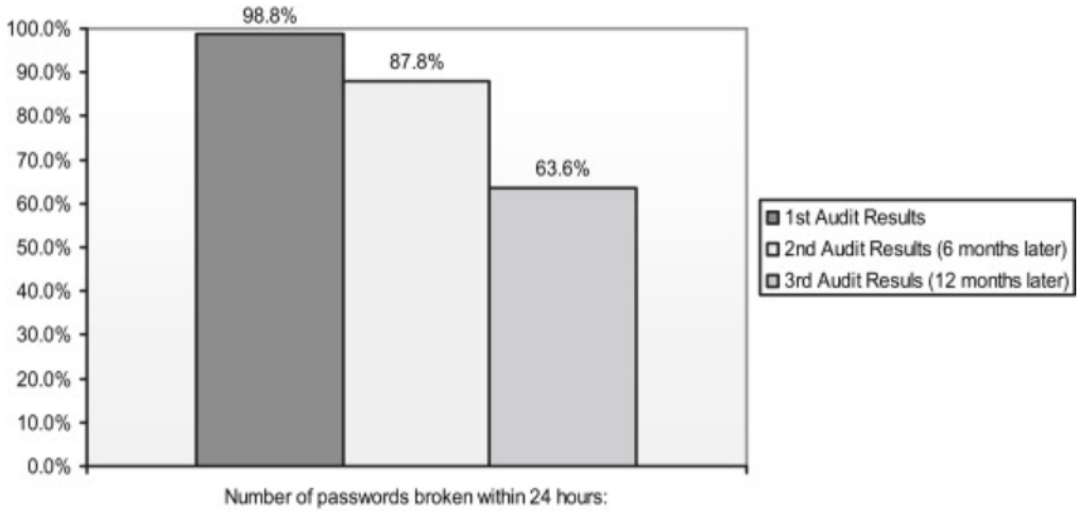
Previous Studies

1. “The positive outcomes of information security awareness training in companies – A case study”
2. “Assessment of the Impact of Information Security Awareness Training Methods on Knowledge, Attitude, and Behavior”
3. “Effects on employees' information security abilities by e-learning”
4. “Improving information security awareness and behavior through dialogue, participation and collective reflection. An intervention study”
5. “A Descriptive Review and Classification of Organizational Information Security Awareness Research”

“The positive outcomes of information security awareness training in companies – A case study”

- Yearlong study on 2900 employees
- Test ISA training success based on password usage, password quality, compliance of password policies
- 3 password audits were done
- Results
  - Employes chose stronger passwords over time
  - Greater tendency to comply with policies
  - Suggests there should be ongoing ISA campaigns





“Assessment of the Impact of Information Security Awareness Training Methods on Knowledge, Attitude, and Behavior”

- Test the impact of multiple ISA strategies on knowledge, attitude, and behavior
- Measured effectiveness with HAIS-Q (Human Aspects of Information Security Questionnaire) – 140 participants
- Knowledge - comparable in improvement across strategies
  - Attitude - improved attitude outside of lecture only
  - Behavior - lecture + game outperformed other strategies

Group	Treatment	Participants	Male	Female
A	Lecture + Video	33	15	18
B	Lecture + Reading	32	18	14
C	Lecture + Game	32	16	16
D	Lecture Only	31	14	17
	TOTAL	128	63	65

“Effects on employees' information security abilities by e-learning”

- Pre and post-assessment of knowledge and behavior on 1900 employees
- Measuring indexes:
  - Knowledge – definitions of risk, security policy, integrity, and physical security
  - Awareness – security versus functionality, reporting willingness, and importance of generic security and safety means
  - Behavior – written passwords, locking PC when leaving it, reporting incidents upon detection
- Results showed improvements across all three areas

“Improving information security awareness and behavior through dialogue, participation and collective reflection. An intervention study”

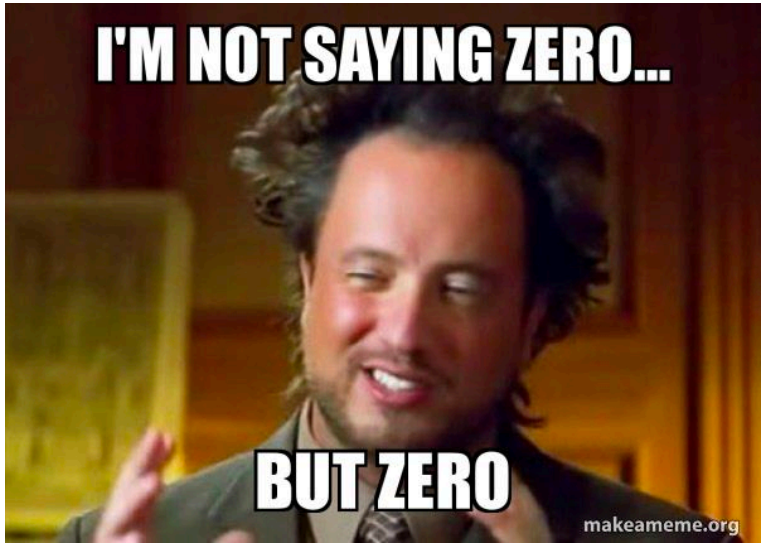
- 6 workshops involving 197 employees
- Each workshop encouraged group dialogue and tested them against common scenarios
- 1 survey before, 2 surveys after
- Survey measured indexes on awareness and behavior
  - Very limiting and outdated
- 2/3 reported “yes” to perceived change of awareness and behavior
- Indicates that group-based discussion is more effective than lectures

“A Descriptive Review and Classification of Organizational Information Security Awareness Research”

- Purpose – review and classify current ISA knowledge
- Sample 59 peer-reviewed articles from 2008-2018
- Results = ISA is evolving and behavioral research needs explored

Q2 - Do current ISA trainings incorporate cognitive psychology?

Previous Studies



Method

- Manually gathered a list of top 50 most popular ISA trainings
- Narrowed to 39
- Determine key words
- Web scrape
- Determine threshold
- Analyze

	A	B
1	**Training Program**	**URL**
2	KnowBe4	https://www.knowbe4.com/en/knowbe4-training-modules-overview/
3	HoxHunt	https://www.hoxhunt.com/product/security-awareness-training
4	MetaComplinace	https://www.metacompliance.com/security-awareness-training
5	Infosec IQ	https://www.infosecinstitute.com/iq/awareness/
6	Proofpoint	https://www.proofpoint.com/sites/default/files/pfpt-us-sb-security-awareness-training-content.pdf
7	PhishingBox	https://www.phishingbox.com/solutions/security-awareness-training
8	Artic Wolf	https://arcticwolf.com/solutions/managed-security-awareness/
9	SoSafe	https://sosafe-awareness.com/
10	Ninjio	https://ninjio.com/
11	Immersive labs	https://www.immersivelabs.com/
12	Cyberready	https://cybeready.com/cab
13	Usecure	https://www.usecure.io/security-awareness-training
14	Curricula	https://www.curricula.com/security-awareness-training-topics
15	Webroot	https://www.webroot.com/us/en/business/products/security-awareness-training
16	Hook Security	https://www.hooksecurity.co/solutions/security-awareness-training
17	CyberHoot	https://cyberhoot.com/features/
18	Phished	Cyber Resilience | Change behaviour, prevent cyber incidents. (phished.io)
19	Defenidify	https://www.defendify.com/layered-security/policies-training/cybersecurity-awareness-training/
20	ThreatCop	https://threatcop.com/threatcop-security-awareness-training
21	CyberVista	https://www.cybervista.net/resolve/cybersecurity-awareness/
22	Insipired Elearning	https://inspiredelearning.com/course-catalog/
23	Avatao	https://avatao.com/security-awareness/
24	global learning systems	https://globallearningsystems.com/security-awareness-training/
25	Symantec	https://vox.veritas.com/legacyfs/online/veritasdata/SSAP_5_Datasheet.pdf
26	CultureAI	Security awareness coaching (culture.ai)
27	Haekka	Training Embedded Into Slack | Haekka
28	ThreatAdvice	ThreatAdvice Cybersecurity Education
29	SANS	Security Awareness Training | SANS Security Awareness
30	Barracuda	Security Awareness Training | Barracuda Networks
31	Mimecast	Security Awareness Training | Awareness Training | Mimecast
32	Proofpoint	Security Awareness Training - Cybersecurity Education Online | Proofpoint US
33	living security	Advanced Security Awareness Training | Living Security
34	Boxphish	Best Cyber Security Awareness Training | Boxphish
35	HacWare	Security Awareness Training & Learning Management System (hacware.com)
36	confense phishme	Cofense Security Awareness Training - Cofense
37	cybercoach	CyberCoach Safety & Security Awareness Training
38	CIRA	CIRA Cybersecurity Awareness Training
39	Security Mentor	Security Awareness Training | Security Mentor, Inc.
40	Wizer	Cyber Security Awareness Training for Employees | Wizer (wizer-training.com)
41	CybSafe	Why CybSafe? Security Awareness and Training Solutions
42	Dcoya	DCOYA Behave - DCOYA - Making Cybersecurity Personal
43	Elevate Secuity	Elevate Engage - Human Risk Monitoring and Management - Elevate Security
44	Riot	Grow the cybersecurity culture in your team · Riot (tryriot.com)

Keywords

1. Psychology

MATERIAL KEYWORDS

2. Behavior
3. Memory
4. Attention
5. Perception

DESIGN KEYWORDS

6. Engaging
7. Interactive
8. Scenario
9. Concise

The magic

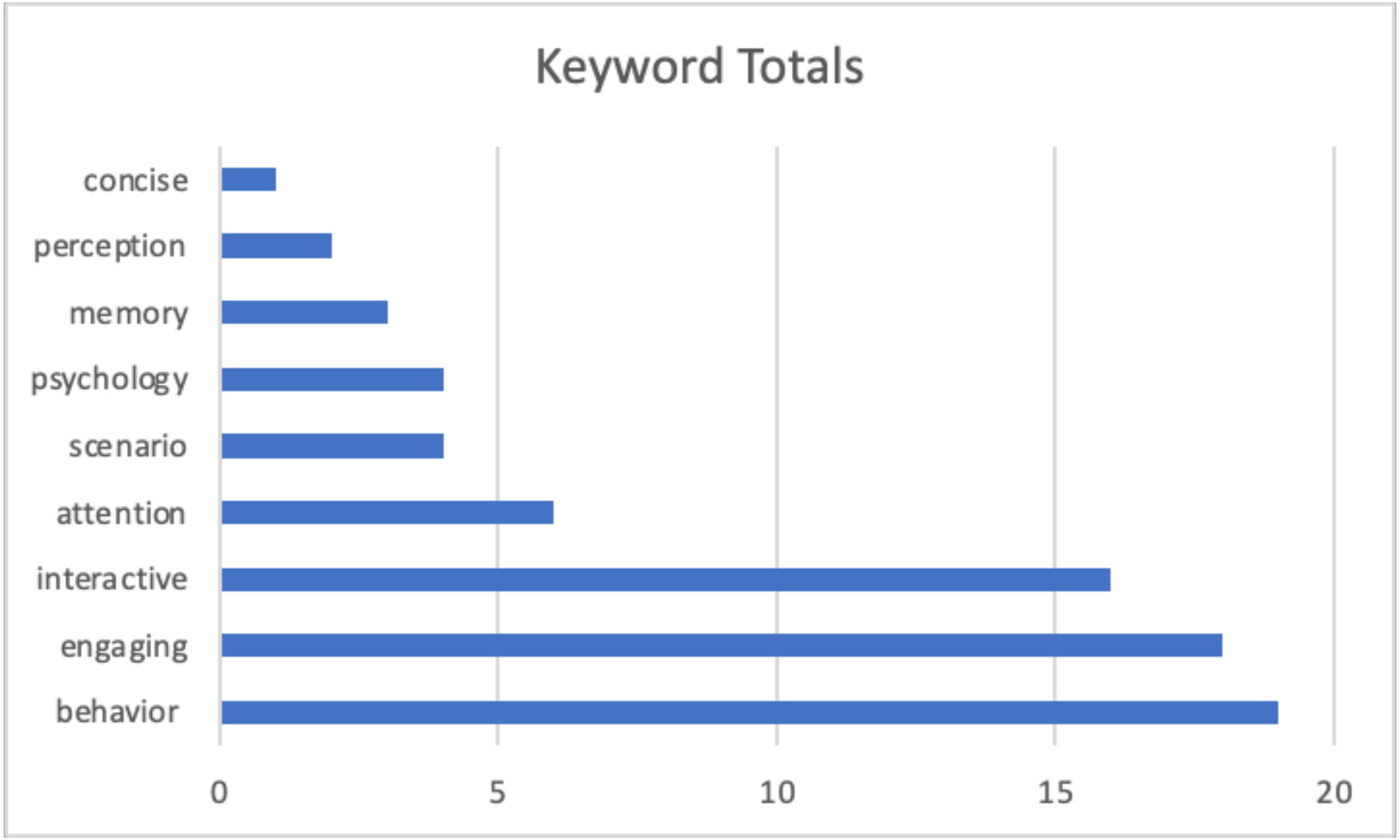
```
# List of keywords to search for
keywords <- c("psychology", "behavior", "memory", "attention", "perception","engaging", "scenario", "interactive", "concise")

# Create a data frame to store the results
results <- data.frame(links = character(length(websites)),
                      keyword1 = integer(length(websites)),
                      keyword2 = integer(length(websites)),
                      keyword3 = integer(length(websites)),
                      keyword4 = integer(length(websites)),
                      keyword5 = integer(length(websites)),
                      keyword6 = integer(length(websites)),
                      keyword7 = integer(length(websites)),
                      keyword8 = integer(length(websites)),
                      keyword9 = integer(length(websites)),
                      stringsAsFactors = FALSE)

# Loop through each website in the list
for (i in 1:length(websites)) {

  # Scrape the website using rvest
  page <- read_html(websites[i])

  # Search for each keyword and update the results data frame accordingly
  results[i, "links"] <- websites[i]
  for (j in 1:length(keywords)) {
    results[i, paste0("keyword", j)] <- as.numeric(any(str_detect(page %>% html_text(), keywords[j])))
  }

  # fill in any remaining columns with NA
  results[i, is.na(results[i, ])] <- NA
}

# Write the results to a CSV file
write.csv(results, "SearchResults.csv", row.names = FALSE)

```

Testing...

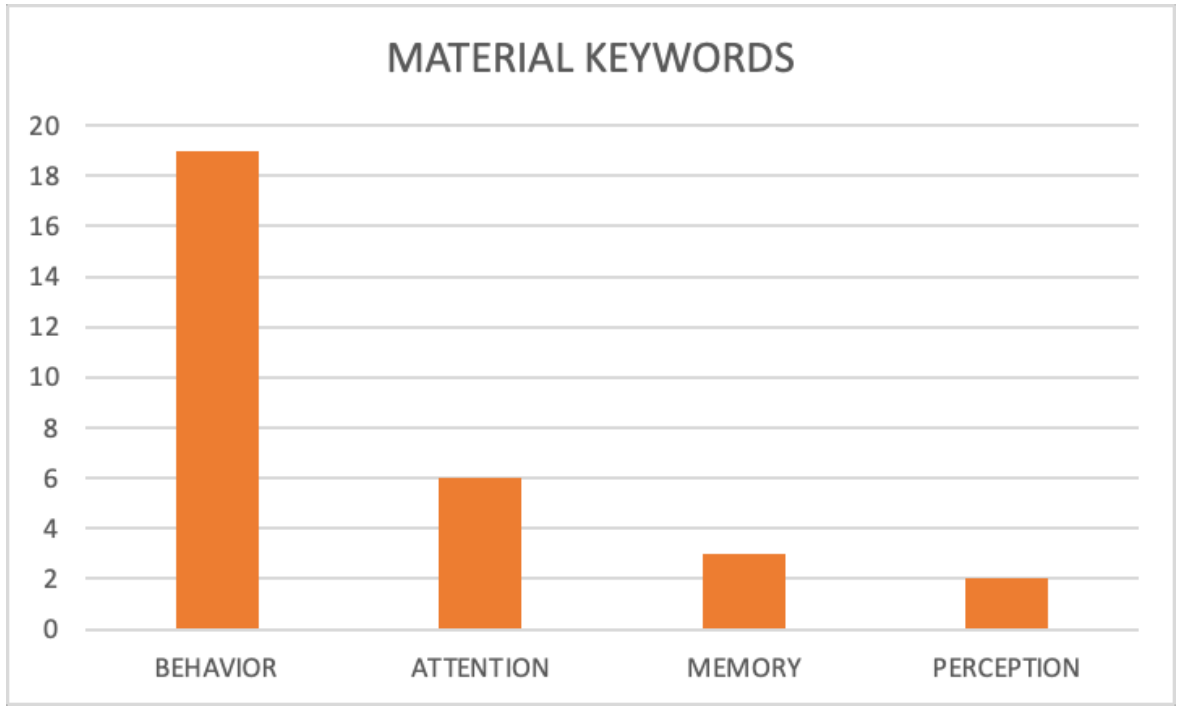
Links	Security
https://www.knowbe4.com/en/knowbe4-training-modules-overview/	1
https://www.hoxhunt.com/product/security-awareness-training	1
https://www.metacompliance.com/security-awareness-training	1
https://www.infosecinstitute.com/iq/awareness/	1
https://www.phishingbox.com/solutions/security-awareness-training	1
https://arcticwolf.com/solutions/managed-security-awareness/	1
https://sosafe-awareness.com/	1
https://ninjio.com/	1
https://www.immersivelabs.com/	1
https://cybeready.com/cab	1
https://www.usecure.io/security-awareness-training	1
https://www.curricula.com/security-awareness-training-topics	1
https://www.webroot.com/us/en/business/products/security-awareness-training	1
https://www.hooksecurity.co/solutions/security-awareness-training	1
https://cyberhoot.com/features/	1
https://www.defendify.com/layered-security/policies-training/cybersecurity-awareness-training/	1
https://threatcop.com/threatcop-security-awareness-training	1
https://www.cybervista.net/resolve/cybersecurity-awareness/	1
https://inspiredelearning.com/course-catalog/	1
https://avatao.com/security-awareness/	1
https://globallearningsystems.com/security-awareness-training/	1
https://www.culture.ai/platform/security-awareness-coaching	1
https://www.haekka.com/training	1
https://www.threatadvice.com/educate	1
https://avatao.com/security-awareness/	1
https://www.barracuda.com/products/email-protection/security-awareness-training	1
https://www.proofpoint.com/us/products/security-awareness-training	1
https://www.livingsecurity.com/solutions/advanced-enterprise-security-awareness-training	1
https://www.boxphish.com/cyber-security-awareness-training/	1
https://www.hacware.com/train.html	1
https://cofense.com/knowledge-center/security-awareness-training/	1
https://www.cybercoach.com/cybercoach-safety-security-awareness-training	1
https://www.cira.ca/cira-cybersecurity-services/cybersecurity-awareness-training	1
https://www.securitymentor.com/products-services/security-awareness-training	1
https://www.wizer-training.com/cyber-security-awareness-training	1
https://www.cybsafe.com/why-cybsafe/	1
https://dcoya.com/dcoya-behave/	1
https://elevatesecurity.com/solutions/engage/	1
https://tryriot.com/albert	1

Web Scraping Results

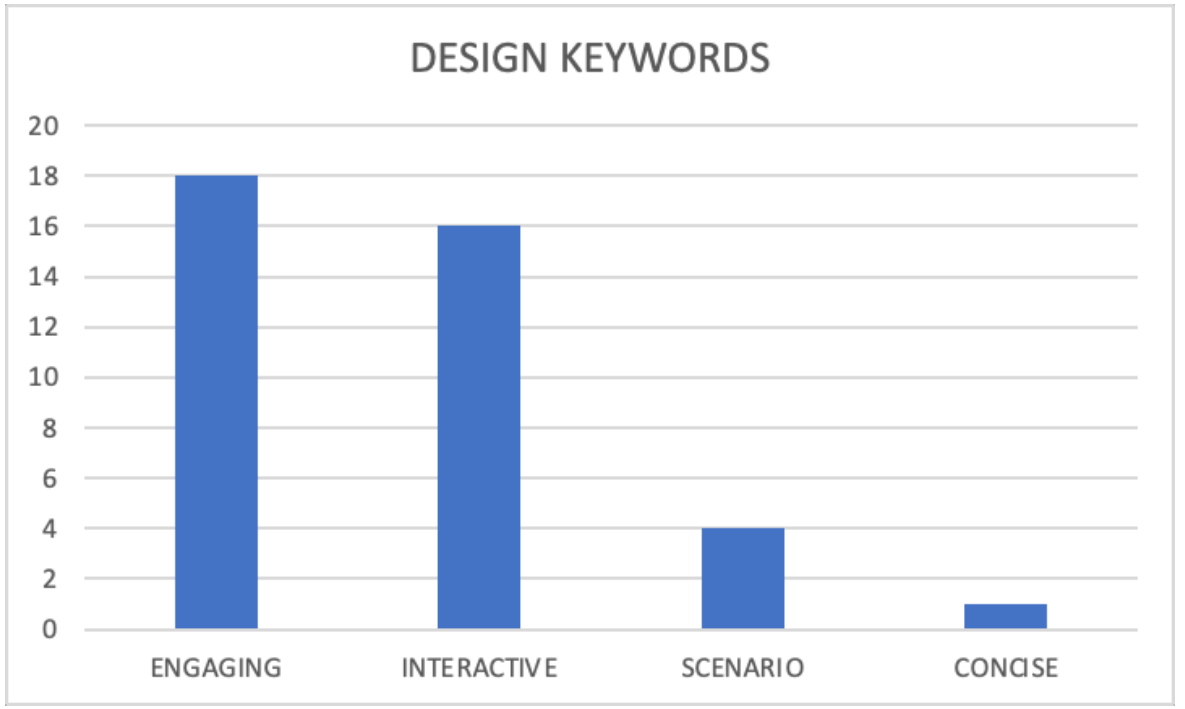
Links	psychology	behavior	memory	attention	perception	engaging	scenario	interactive	concise	Totals
https://www.knowbe4.com/en/knov	0	1	1	0	0	1	1	1	0	5
https://inspiredelearning.com/cours	0	1	1	1	0	0	0	1	0	4
https://www.securitymentor.com/p	0	1	0	1	0	1	0	1	0	4
https://www.wizer-training.com/cyb	1	0	0	1	0	1	0	1	0	4
https://sosafe-awareness.com/	0	1	0	0	0	1	0	1	0	3
https://ninjio.com/	0	1	0	1	0	1	0	0	0	3
https://www.immersivelabs.com/	0	0	0	0	1	0	1	1	0	3
https://cofense.com/knowledge-cent	0	0	0	0	0	1	0	1	1	3
https://www.cybsafe.com/why-cybsa	1	1	0	0	1	0	0	0	0	3
https://www.hoxhunt.com/product/	0	1	0	0	0	1	0	0	0	2
https://www.metacompliance.com/	0	0	0	1	0	1	0	0	0	2
https://www.infosecinstitute.com/ic	0	1	0	0	0	0	0	1	0	2
https://arcticwolf.com/solutions/ma	0	1	0	0	0	0	0	1	0	2
https://cybeready.com/cab	0	0	0	0	0	1	0	1	0	2
https://www.usecure.io/security-aw	0	0	0	0	0	1	0	1	0	2
https://www.webroot.com/us/en/bu	0	1	0	0	0	0	0	1	0	2
https://www.hooksecurity.co/soluti	0	0	0	0	0	1	0	1	0	2
https://www.defendify.com/layered-	0	0	0	0	0	1	0	1	0	2
https://threatcop.com/threatcop-se	0	0	0	0	0	1	0	1	0	2
https://globallearningsystems.com/s	0	1	0	1	0	0	0	0	0	2
https://www.culture.ai/platform/sec	0	0	1	0	0	1	0	0	0	2
https://www.barracuda.com/produc	0	1	0	0	0	1	0	0	0	2
https://www.livingsecurity.com/solu	0	1	0	0	0	1	0	0	0	2
https://dcoya.com/dcoya-behave/	1	1	0	0	0	0	0	0	0	2
https://elevatesecurity.com/solution	1	1	0	0	0	0	0	0	0	2
https://www.curricula.com/security	0	1	0	0	0	0	0	0	0	1
https://www.cybervista.net/resolve/	0	1	0	0	0	0	0	0	0	1
https://www.haekka.com/training	0	0	0	0	0	0	1	0	0	1
https://www.proofpoint.com/us/pro	0	1	0	0	0	0	0	0	0	1
https://www.boxphish.com/cyber-se	0	0	0	0	0	1	0	0	0	1
https://www.hacware.com/train.htr	0	0	0	0	0	0	0	1	0	1
https://www.cybercoach.com/cyber	0	0	0	0	0	0	1	0	0	1
https://www.cira.ca/cira-cybersecuri	0	0	0	0	0	1	0	0	0	1
https://tryriot.com/albert	0	1	0	0	0	0	0	0	0	1
https://www.phishingbox.com/solut	0	0	0	0	0	0	0	0	0	0
https://cyberhoot.com/features/	0	0	0	0	0	0	0	0	0	0
https://avatao.com/security-awarene	0	0	0	0	0	0	0	0	0	0
https://www.threatadvice.com/educ	0	0	0	0	0	0	0	0	0	0
https://avatao.com/security-awarene	0	0	0	0	0	0	0	0	0	0


Keyword Totals
concise
perception
memory
psychology
scenario
attention
interactive
engaging
behavior
0
5
10
15
20

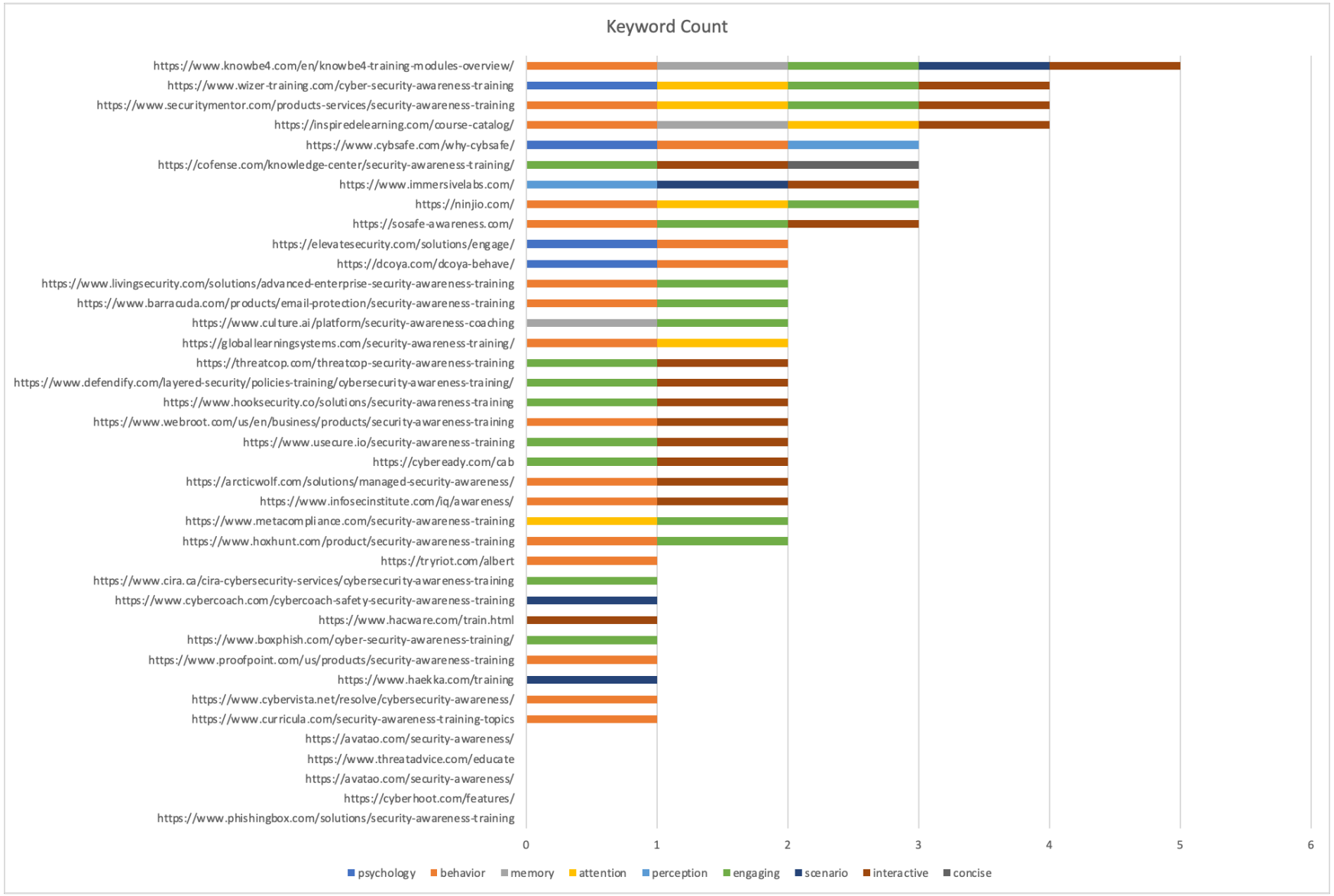
Material (43%)



Design (57%)



Could indicate that more ISA trainings include cognitive psychology in design rather than in material?



Keywords	Count
5	1 (2%)
4	3 (8%)
3	5 (13%)
2	16 (41%)
1	9 (23%)
0	5 (13%)

Hypothesis

If the ISA shares two or more keywords across **material and design** categories, then the ISA effectively incorporates cognitive psychology.

The more shared keywords present, the greater the representation of cognitive psychology in the ISA.

The greater representation, the better the ISA training program.

Results

- 5% of ISA training programs meet our hypothesis.
- Knowbe4 and Security Mentor are the two best ISA training programs.

Links	MATERIAL KEYWORD COUNT	DESIGN KEYWORD COUNT	SHARED TOTAL	%
https://www.knowbe4.com/en/knov	2	3	5	5%
https://www.securitymentor.com/pr	2	2	4	
https://inspiredelearning.com/cours	3	1	4	
https://sosafe-awareness.com/	1	2	3	
https://ninjio.com/	2	1	3	13%
https://www.immersivelabs.com/	1	2	3	
https://www.wizer-training.com/cyb	1	2	3	
https://www.hoxhunt.com/product/s	1	1	2	
https://www.metacompliance.com/s	1	1	2	
https://www.infosecinstitute.com/ic	1	1	2	
https://arcticwolf.com/solutions/ma	1	1	2	20%
https://www.webroot.com/us/en/bu	1	1	2	
https://www.culture.ai/platform/sec	1	1	2	
https://www.barracuda.com/produc	1	1	2	
https://www.livingsecurity.com/solu	1	1	2	
https://www.phishingbox.com/solut	0	0	0	
https://cyberhoot.com/features/	0	0	0	
https://avatao.com/security-awaren	0	0	0	
https://www.threatadvice.com/educ	0	0	0	
https://avatao.com/security-awaren	0	0	0	
https://www.curricula.com/security-	1	0	0	
https://www.cybervista.net/resolve/	1	0	0	
https://www.haekka.com/training	0	1	0	
https://www.proofpoint.com/us/pro	1	0	0	
https://www.boxphish.com/cyber-se	0	1	0	
https://www.hacware.com/train.htn	0	1	0	
https://www.cybercoach.com/cyberc	0	1	0	62%
https://www.cira.ca/cira-cybersecur	0	1	0	
https://tryriot.com/albert	1	0	0	
https://cybeready.com/cab	0	2	0	
https://www.usecure.io/security-aw	0	2	0	
https://www.hooksecurity.co/solutio	0	2	0	
https://www.defendify.com/layered-	0	2	0	
https://threatcop.com/threatcop-sec	0	2	0	
https://globallearningsystems.com/s	2	0	0	
https://dcoya.com/dcoya-behave/	1	0	0	
https://elevatesecurity.com/solution	1	0	0	
https://cofense.com/knowledge-cen	0	3	0	
https://www.cybsafe.com/why-cybs	2	0	0	

Do current ISA trainings incorporate cognitive psychology?

No

Limitations

- Our analysis did not look at the programs in entirety but rather their descriptions
- Not all keywords might be accurate indicators
- The threshold we determined may not reflect the degree of ISA that incorporates cognitive psychology by material and design

Q3 - How can we more effectively design ISA training to include cognitive psychology to achieve greater awareness?

1. Changing design
2. Changing material

Use of fake logos to build trust

Can you spot the real Office 365 Login page?!?

Design

- Attention ~ draw and keep attention
- Memory ~ short term- long-term memory
- Perception ~ Simplify
- Motivation ~ "fear!"
- Decision Making ~ provide guidelines, practice
- Feedback ~ Timely feedback

Design

- Goal: allow for optimal formation of schemas to improve information recall
  - Interactive – require communication and team work
  - Engaging – scenarios / immersive simulated environments
  - Continual – monthly basis
  - Concise – 30 min per week

Schema

What is schema and how does it relate ?

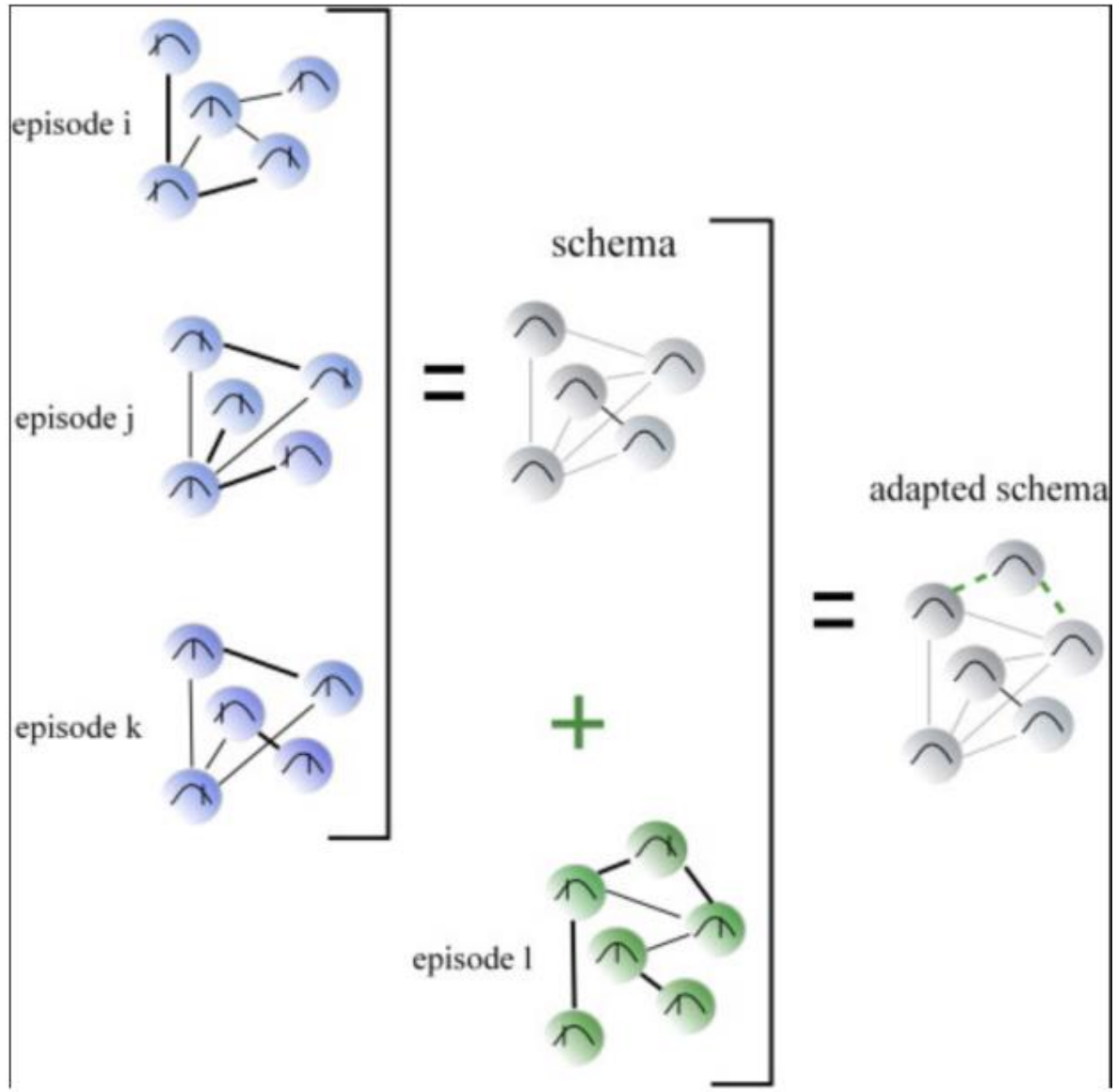
V. E. Ghosh and A. Gilboa, "What is a memory schema? A historical perspective on current neuroscience literature," Neuropsychologia, vol. 53, pp. 104–114, 2014.

"any pattern of relationships among data stored in memory" - Heuer

"complex construct that is employed to influence the encoding and retrieval of episodic memory" - Heuer

Schema..

- Muscle memory of acceptable safe behavior
- Guiding behavior, encoding of new information, inferential elaboration, expedited retrieval process
- So ingesting multiple episodes of a concept creates a schema and once a schema has been built, we can input additional episodes and concepts to modify and create an adapted schema
- Associative Network Structure, Basis of multiple episodes, Lack of unit detail, Adaptability


episode i
episode j
episode k
schema
=
+
episode l
=
adapted schema

Material

- Attention and Perception ~ Improve attention and perception skills
- Memory ~ Retrieve the "right" information when you need it
- Decision-Making ~ overcoming biases, rational decision making
- Social Influence ~ social cues

Concluding Thoughts

- Measuring effectiveness of ISA is a difficult task
  - Is self-reporting accurate?
- We assessed current ISA trainings on the basis of psychology in both design and material
- Based off research, we made suggestions of how ISA can improve

Lessons Learned

- We had to be creative - there was no template or guiding research
- We made speculations and our ideas have not been tested
  - Ex: How do we know keywords we used could be indicators of anything?
  - Ex: Is there a more accurate way of forming conclusions?
- This is the first step in a larger project that would be unique research