How to Combat the Fake News and Disinformation

Being Used to Attack Your Organization

Stu Sjouwerman
Founder and CEO
KnowBe4, Inc.

Perry Carpenter
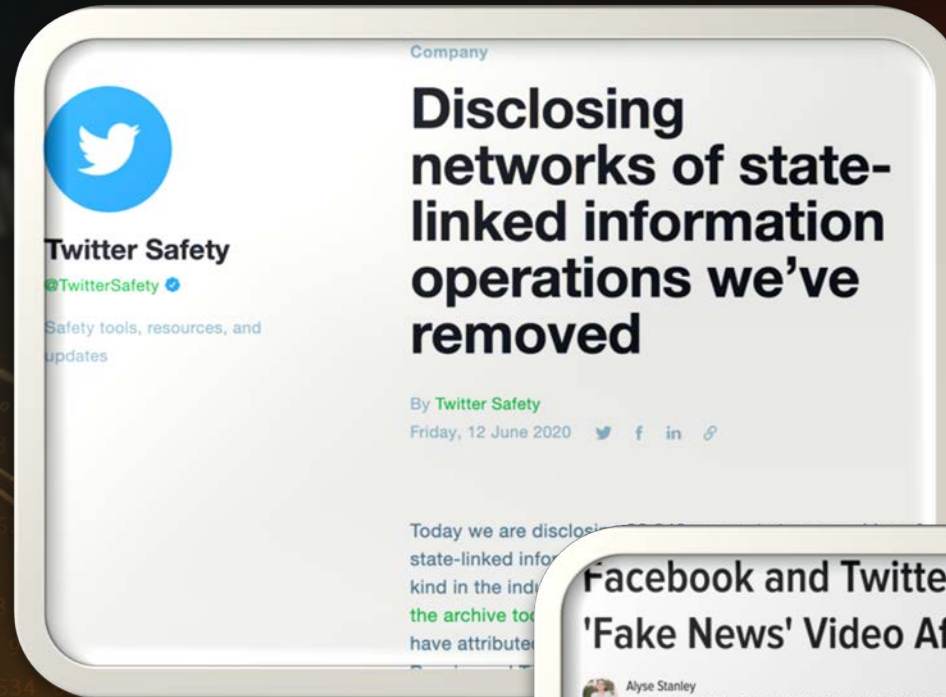Chief Evangelist & Strategy Officer
KnowBe4, Inc.

KnowBe4
Human error. Conquered.

Stu Sjouwerman
Founder and CEO
KnowBe4, Inc.

Perry Carpenter
Chief Evangelist & Strategy Officer
KnowBe4, Inc.

KnowBe4
Human error. Conquered.

Fake News and Influence Operations are Consistently in the News

"All warfare is based on deception."

- Sun Tzu, *The Art of War*

Defining Terms

-- A Dictionary of Disinformation --

FAKE NEWS

dezinformatsia

Disinformation

Misinformation
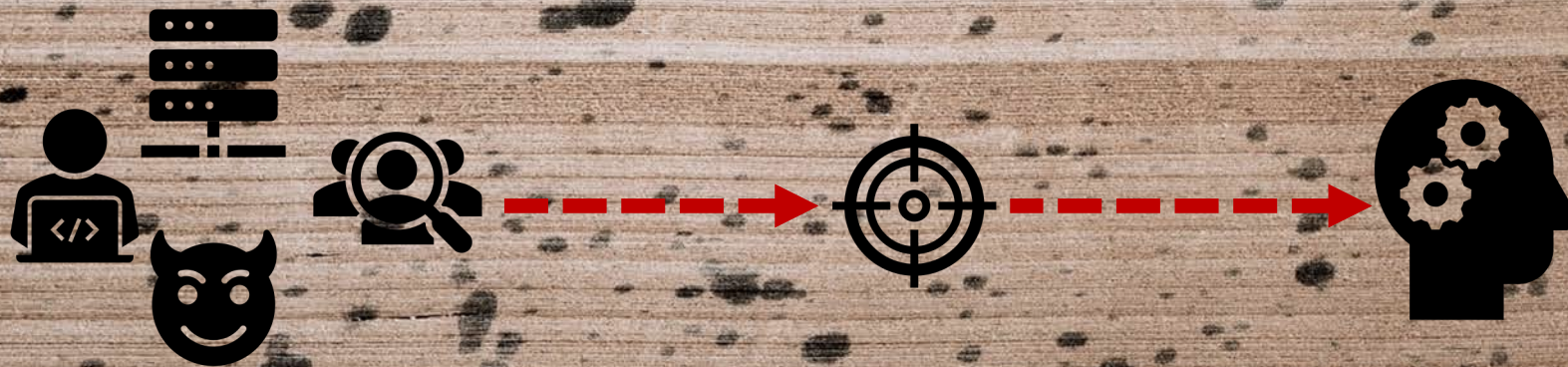
Information Operations
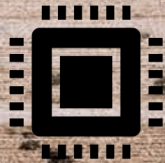
Astroturfing

Sockpuppet
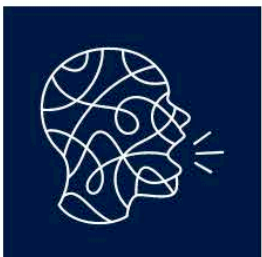
Gaslighting

Fakes

Deepfakes

Cheap-fakes

Generative Adversarial Networks (GANs)

Audio Fakes

The State of Play

- "Fake news" is a politicized term
- We are in a weaponized information situation
- Media is in the *attention* business, resulting in fake news laundering
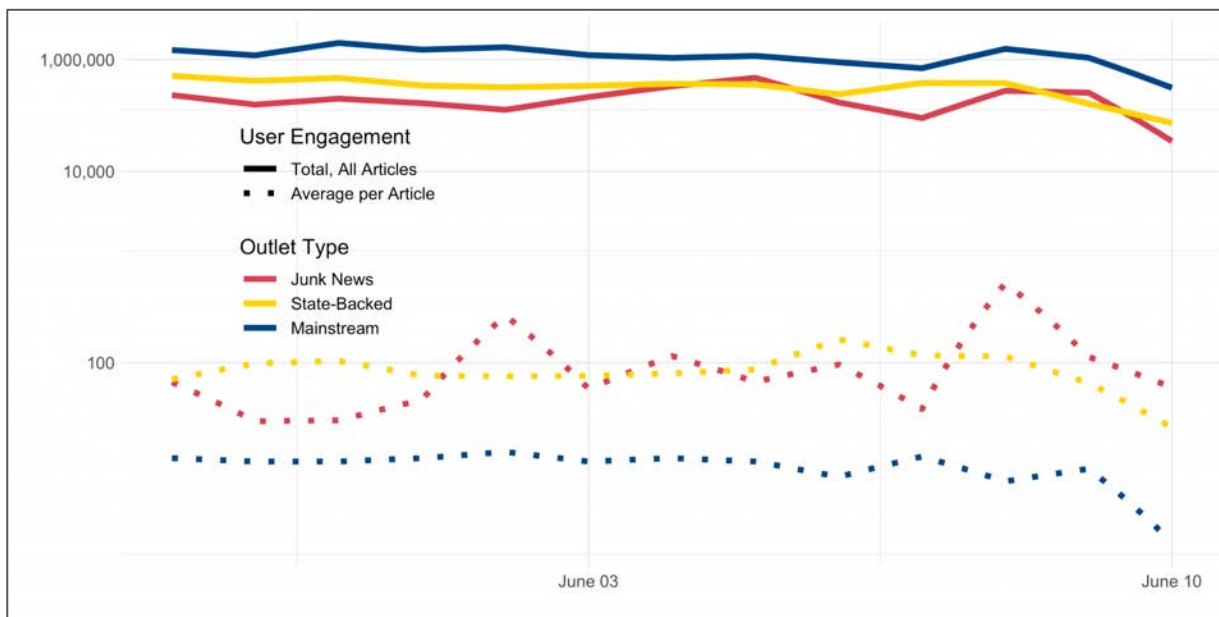- Social media are the most powerful tools on the planet

Coronavirus Misinformation: Weekly Briefings

Engagement Trends – Last 14 Days. Based on authors' calculations using data collected 03/06/2020-10/06/2020.

Source: https://comprop.oii.ox.ac.uk/research/coronavirus-weekly-briefings/

For the week leading up to 15-06-2020 we find:

- The social media distribution network of all coronavirus articles from the top fifteen mainstream news outlets reached over three billion social media users this week, achieving much greater distribution than state-backed and junk health news sources. But the average article from state-backed sources reached nearly 8,000 users, while the average article from mainstream sources reached over 4,000 users and the average junk health article reached over 3,000 users.

- Similarly, all content from all mainstream sources gets the largest amount of total user engagement. But on a per article basis, junk health news gets almost over 100 engagements, state-backed content gets over 75 engagements, and average articles from mainstream sources get under 25 engagements.

- In total, 55% of the engagement with non-mainstream information last week was with state-backed content, and 95% of engagement with state-backed content was made up of engagements with Chinese and Russian media content.

- Thematically, junk health news sources (a) sought to undermine trust in public health experts, (b) used the George Floyd protests to fuel misinformation and racist messages, and (c) advanced false claims about President Trump's success in managing unemployment and economic recovery in the US.

Foreign Influence Operations on Social Media

Countries Attributed by Facebook and Twitter for Engaging in Foreign Influence Operations

FIGURE 3 - FOREIGN INFLUENCE OPERATIONS ON SOCIAL MEDIA
COUNTRIES ATTRIBUTED BY FACEBOOK AND TWITTER FOR ENGAGING IN FOREIGN INFLUENCE OPERATIONS

China
Venezuela
India
Saudi Arabia
Iran
Russia
Pakistan
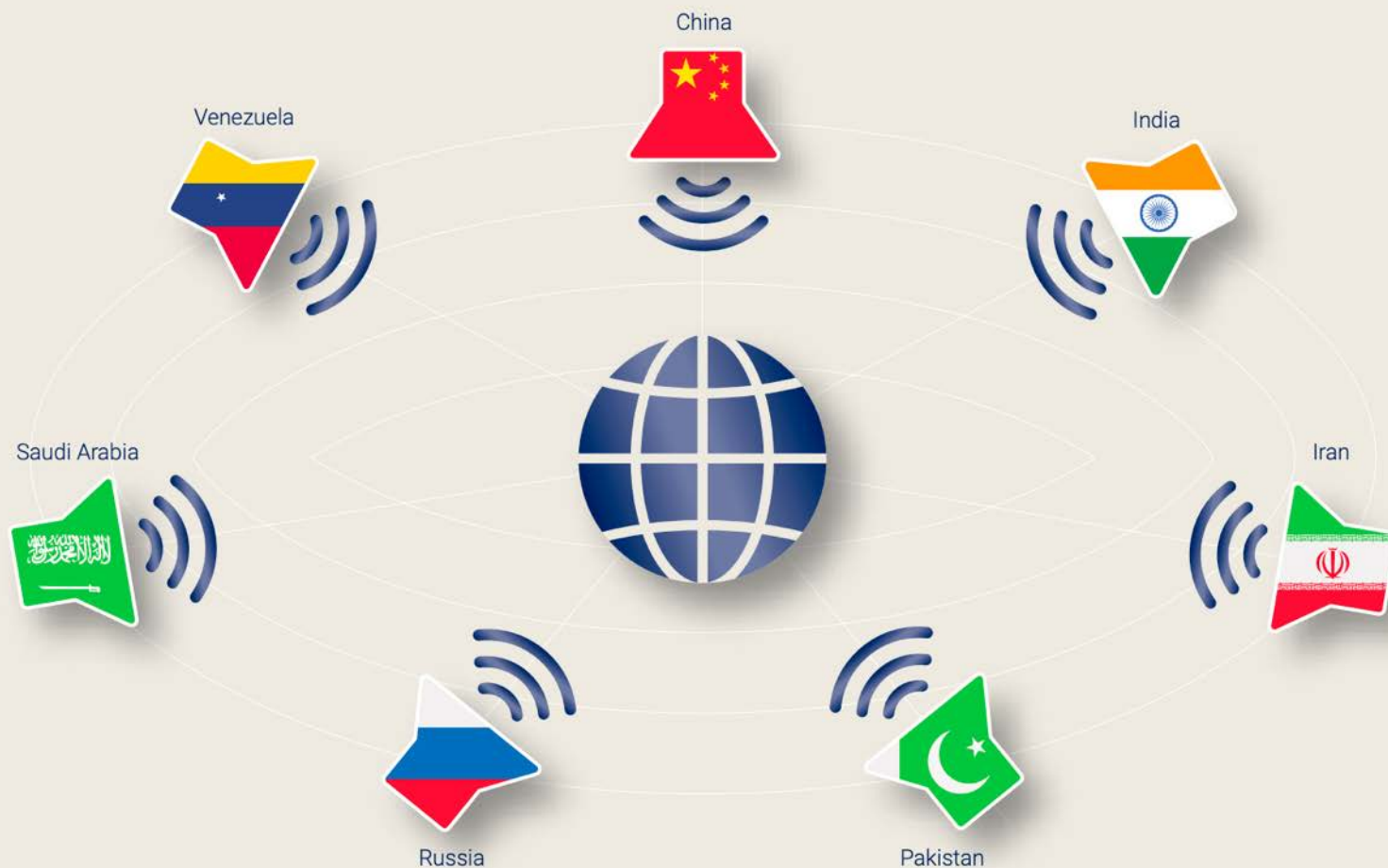
Source: Authors' evaluations based on data collected. **Note:** Facebook has also taken down accounts engaged in 'coordinated inauthentic behaviour' that are not explicitly linked to a government or political party. These takedowns include accounts originating from: Egypt, Macedonia, Kosovo, Thailand, and the United Arab Emirates. Additionally, some cyber troop activity identified by Facebook and Twitter is domestically focused, such as in the case of Bangladesh and Honduras, and is therefore not included in this figure on foreign operations.

Source: The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation
https://www.oii.ox.ac.uk/news/releases/use-of-social-media-to-manipulate-public-opinion-now-a-global-problem-says-new-report/

18

Prominent Platforms for Social Media Manipulation

Social Media Platforms Used for Cyber Troop Activities

Source: The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation
https://www.oii.ox.ac.uk/news/releases/use-of-social-media-to-manipulate-public-opinion-now-a-global-problem-says-new-report/
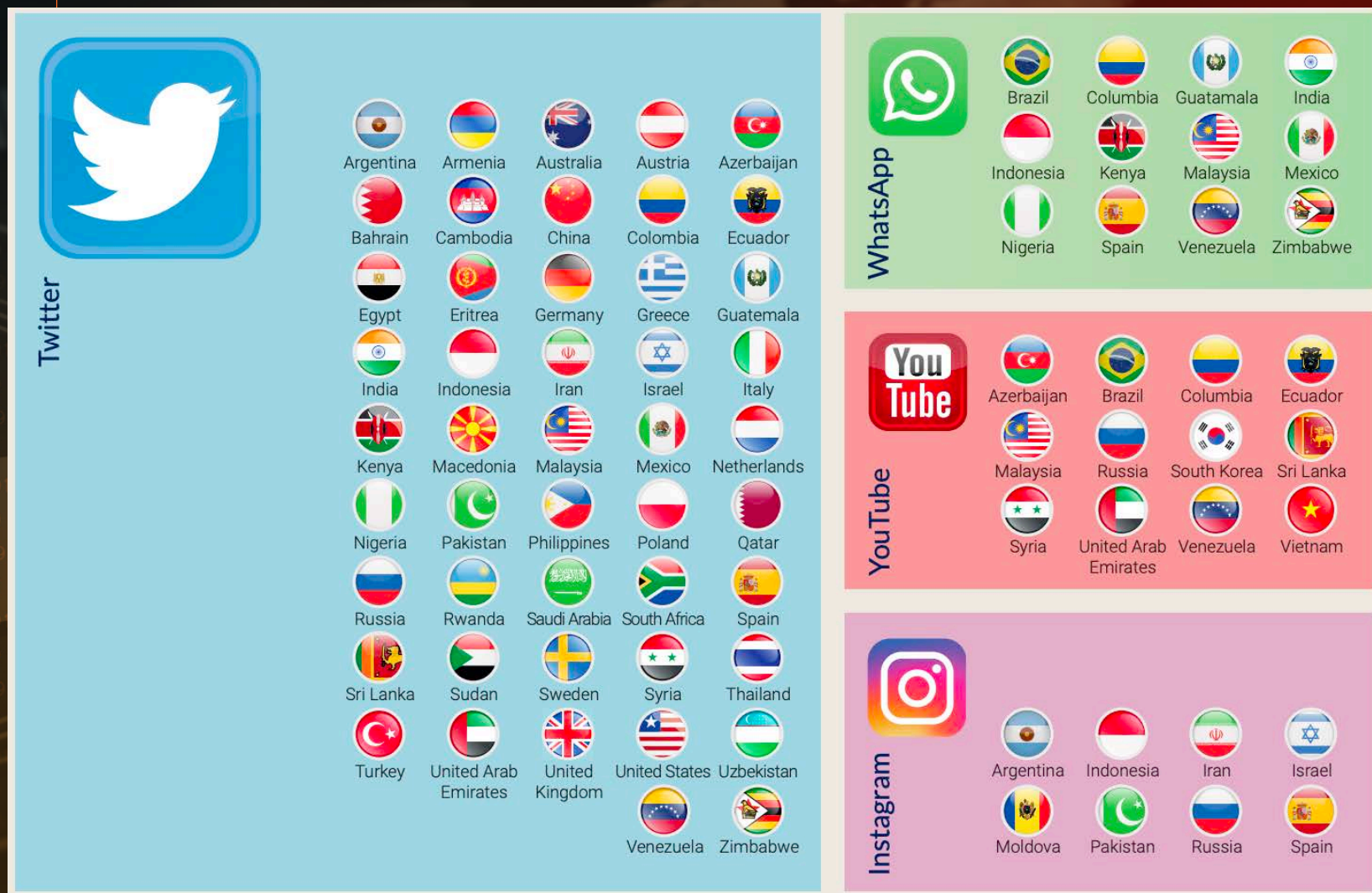
Prominent Platforms for Social Media Manipulation

Social Media Platforms Used for Cyber Troop Activities

Facebook

Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bosnia & Hertc', Brazil, China, Columbia, Croatia, Cuba, Ecuador, Egypt, Eritrea, Ethiopia, Greece, Guatamala, Honduras, Hungary, India, Indonesia, Iran, Israel, Italy, Kenya, Kyrgyzstan, Macedonia, Malaysia, Malta, Mexico, Moldova, Myanmar, Nigeria, Pakistan, Philippines, Russia, South Africa, South Korea, Spain, Sri Lanka, Sudan, Sweden, Syria, Taiwan, Tajikistan, Thailand, Tunisia, Turkey, United Arab Emirates, United Kingdom, United States, Uzbekistan, Venezuela, Vietnam, Zimbabwe
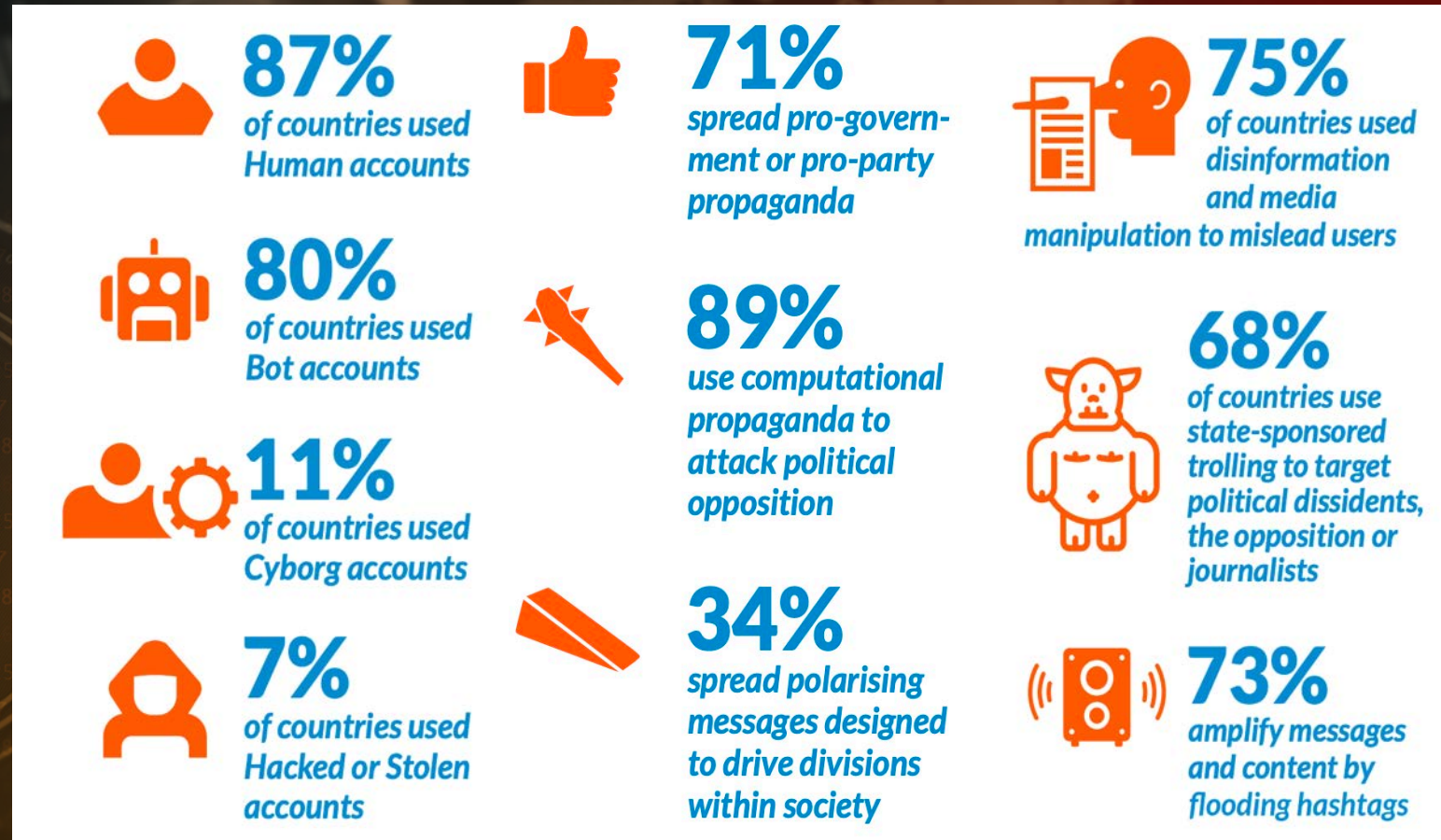
KnowBe4
Human error. Conquered.

Governments Use of Disinformation Campaigns is Widespread

Social Media Platforms Used for Cyber Troop Activities

87% of countries used Human accounts

80% of countries used Bot accounts

11% of countries used Cyborg accounts

7% of countries used Hacked or Stolen accounts

71% spread pro-government or pro-party propaganda

89% use computational propaganda to attack political opposition

34% spread polarising messages designed to drive divisions within society

75% of countries used disinformation and media manipulation to mislead users

68% of countries use state-sponsored trolling to target political dissidents, the opposition or journalists

73% amplify messages and content by flooding hashtags

21

Governments Use of Disinformation Campaigns is Widespread
Social Media Platforms Used for Cyber Troop Activities

Note: this is only a small sample of what is in the full report

Country	Support	Attack Opposition	Distracting	Driving Divisions	Suppressing
China	👍	✈	▼		🔇
Iran	👍	✈	▽		🔇
Russia	👍	✈	▼	▰	🔇
United Kingdom	👍	✈	▼	▰	🔇
United States	👍	✈	▼	▰	🔇

Country	Disinfo'	Mass Reporting	Data-Driven Strategies	Trolls	Amplifying Content
China	✓	✓		✓	✓
Iran	✓	✓	✓	✓	✓
Russia	✓	✓	✓	✓	✓
United Kingdom	✓		✓		✓
United States	✓		✓		✓

Source: The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation
https://www.oii.ox.ac.uk/news/releases/use-of-social-media-to-manipulate-public-opinion-now-a-global-problem-says-new-report/

KnowBe4
Human error. Conquered.

6 Types of Disinformation & Misinformation

1. **Fabricated content**: completely false content.
2. **Manipulated content**: content that includes distortions of genuine information or imagery — a headline, for example, that is made more sensationalist to serve as "clickbait."
3. **Imposter content**: material involving impersonation of genuine sources — by using the branding of an established news agency, for instance.
4. **Misleading content**: information presented in a misleading way — by, for example, presenting comment as fact.
5. **False context of connection**: factually accurate content that is shared with false contextual information — for example, a headline that does not reflect the content of an article.
6. **Satire and parody**: humorous but false stores presented as if they are true. Although this isn't usually categorized as fake news, it may unintentionally fool readers.

House of Commons
Digital, Culture, Media and
Sport Committee

Disinformation and 'fake news': Interim Report

Fifth Report of Session 2017–19

Report, together with formal minutes relating to the report

Ordered by the House of Commons to be printed 24 July 2018

What is the ultimate goal, anyway?

Dismiss an opponents claims or allegations

Distort events to serve political purposes

Distract from one's own activities, and

Dismay those who might otherwise oppose one's goals
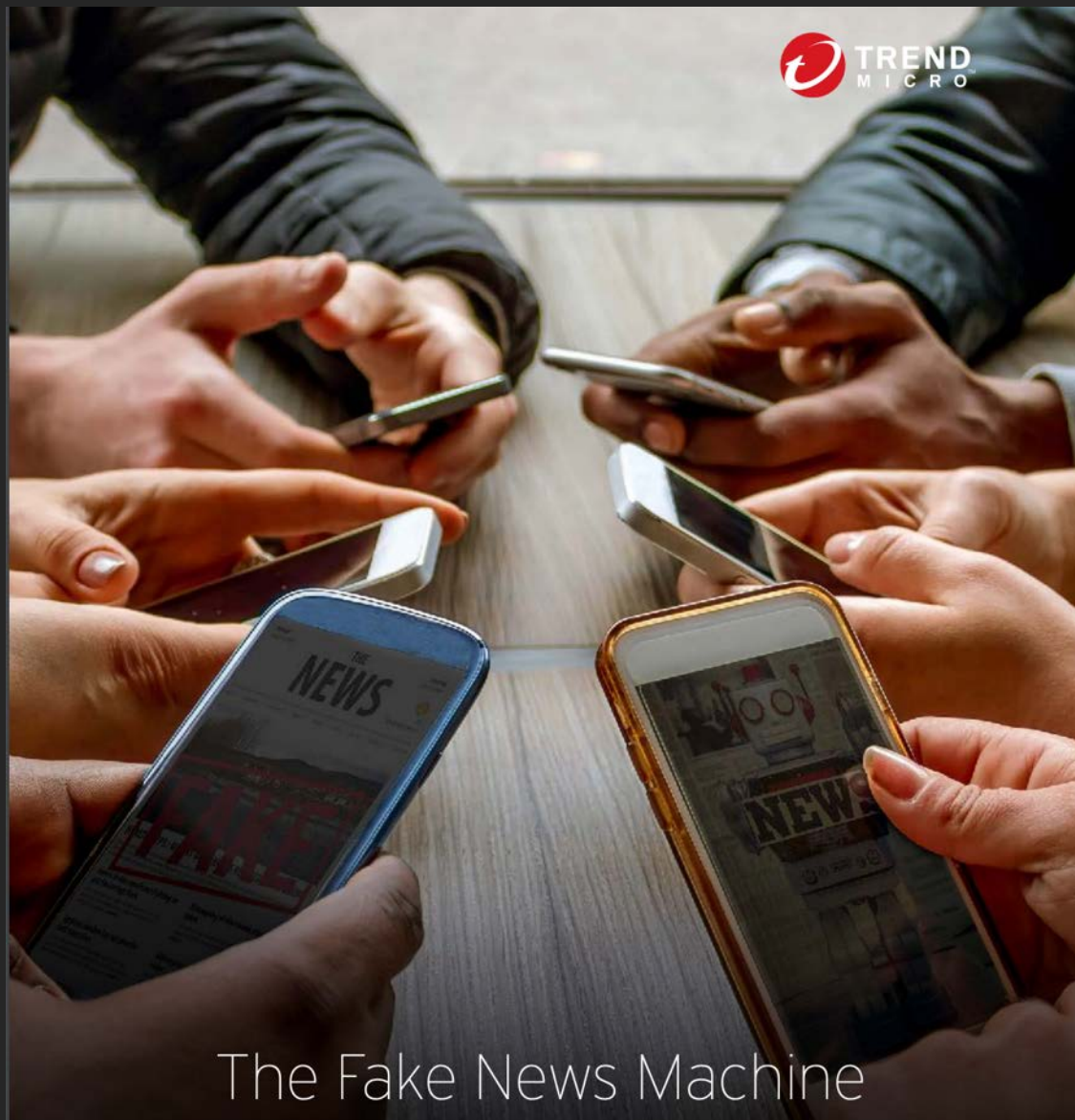
The Fake News Machine
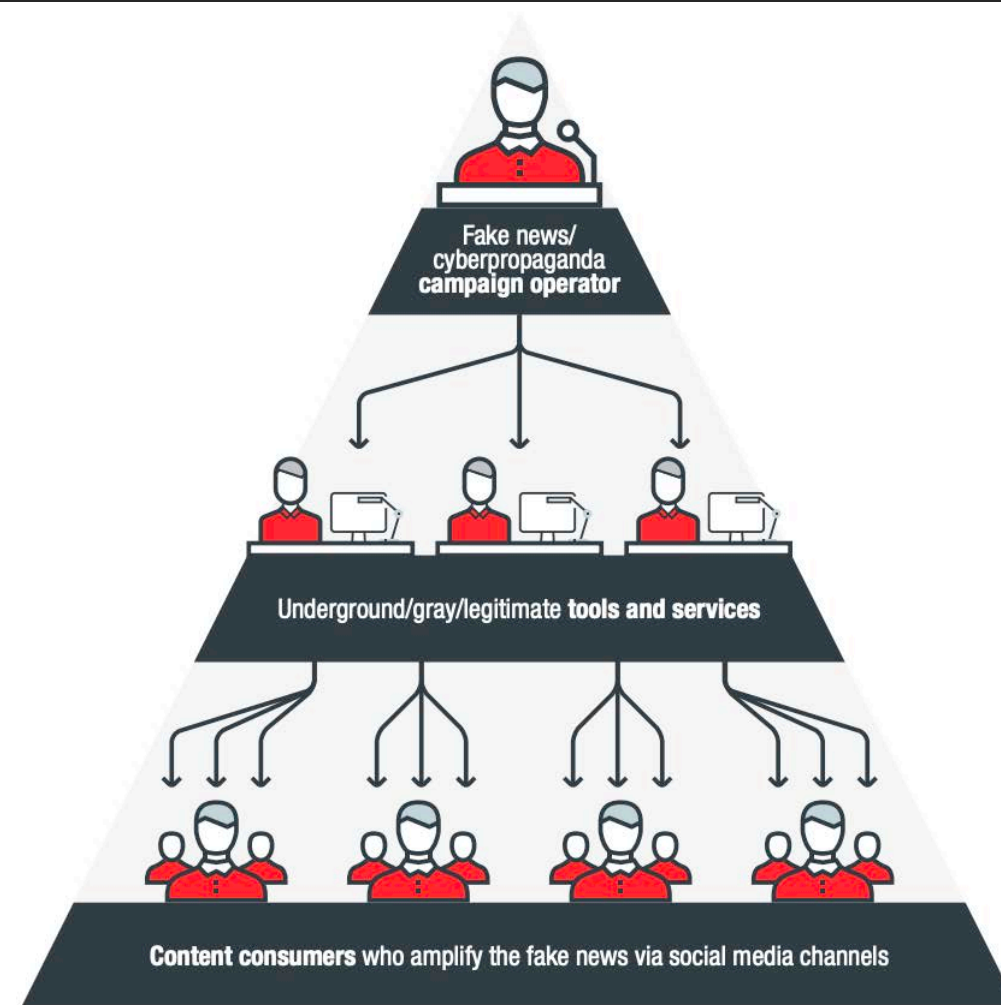
How an operator employs or abuses underground, gray, and legitimate marketplaces to disseminate fake news
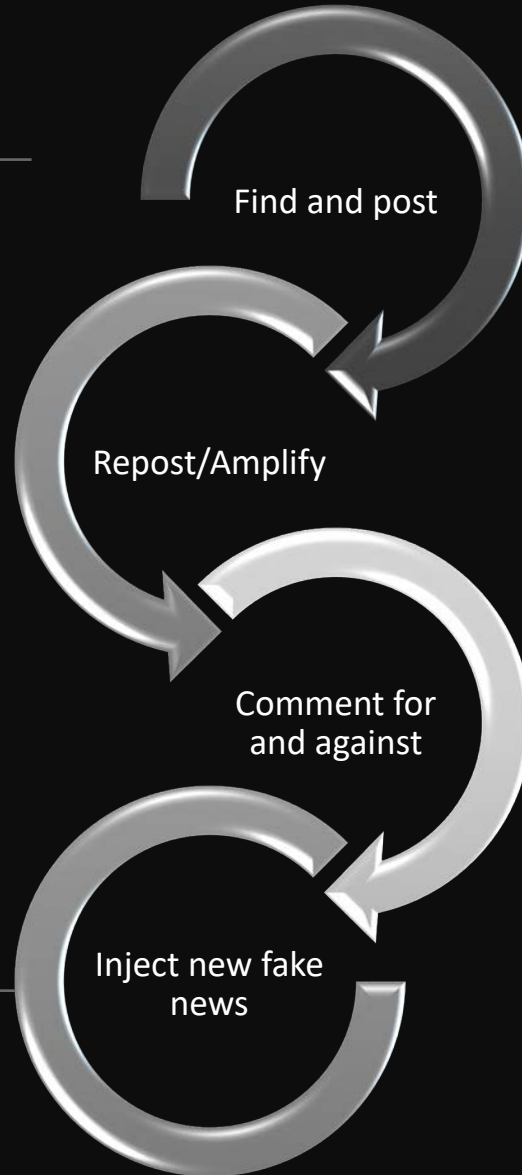
The Fake News Machine How Propagandists Abuse the Internet and Manipulate the Public

Source Actor

Real-time Reenactment

Reenactment Result

Target Actor

New — What is generative media? →

Unique, worry-free model photos

...hance your creative works with photos generated completely by AI. Find model ima...
through our sorted and tagged app, or integrate images via API.

Browse photos

Get API access

Marco Rubio
@marcorubio

Following

British GCHQ that spied on @realDonaldTrump in 2016 is back in the game. Now they intend to use #DeepFakes to support Democrats during midterm elections. No one – especially our closest ally – has the right to threaten our democracy & national security.

9:18 AM - 30 ... 2018

FAKE

16 Retweets 102

15   16   102

ROSEBUD AI   How it Works   Fashion Models   Tech Blog   Pricing   API   LOGIN

None of the models on this page are real.

Get unique AI models and faces for your marketing needs, instantly. No photoshoot needed.

GENERATE YOUR OWN MODEL WITH ROSEBUD AI →

REAL VIDEO   MAY 22   ALTERED VIDEO

Facebook.com/PoliticsWatchDog

CBS THIS MORNING
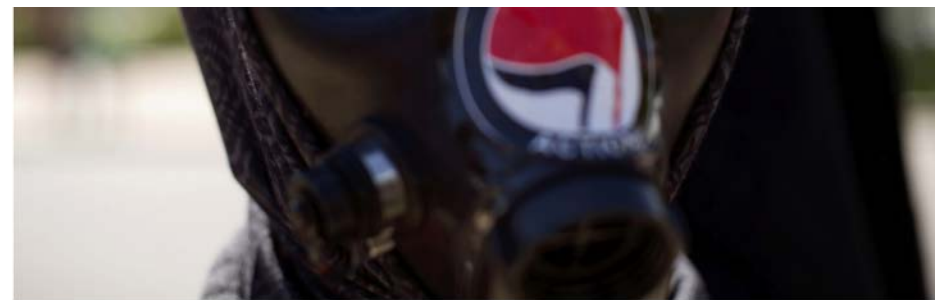
The petition asks the Pentagon to designate antifa groups as terrorists. | Jason Connolly/Getty Images

White House 'antifa' petition written by pro-Trump troll

Nation-states and bad actors create 'truth' – *one meme at a time*

KnowBe4
Human error. Conquered.

Division drives clicks, outrage, disenfranchisement, etc...

Joshua Detar
@JoshDetar
Follow

@HillaryClinton

Don't forget to vote tomorrow! Avoid long lines just text your vote for Hillary. Only one vote per phone! #ImWitHer

Fountainhead
@Fountainhead55
Follow

Hillary voters get to vote via text, but Trump supports have to go to the polls! NOT FAIR!
#imwithher #hillary

Vote Early
Text "Hillary" to 59925
Today
Paid For By Hillary For President

MUST BE 18 YEARS OR OLDER TO VOTE. MUST BE A LEGAL CITIZEN OF THE UNITED STATES. VOTE BY TEXT NOT AVAILABLE IN GUAM, PUERTO RICO, ALASKA, OR HAWAII.

2:10 PM - 7 Nov 2016

Texas Lone Star
@SouthLoneStar
Follow

Muslim woman pays no mind to the terror attack, casually walks by a dying man while checking phone
#PrayForLondon #Westminster #BanIslam

Suggested Page

Defend the 2nd
Sponsored

The community of 2nd Amendment supporters, guns lovers & patriots

DEFEND THE 2ND

Defend the 2nd
Community
96,678 people like this.
Like Page

Heart of Texas shared their event.
Sponsored

Fellow Texans! It's time to say a strong NO to the establishment robbers. It is unacceptable for us to see them ruin all we've been building for decades. For centuries. The establishment thinks they can treat us like stupid sheep but they are wrong. We won't put up with this anymore. The corrupt media does not talk about the crimes committed by Killary Rotten Clinton, neither does it mention the leaked emails but it would rather keep on kicking around some outdated tapes feat... See More

GET READY TO SECEDE!

NOV 5 Get Ready to Secede!
Sat 1 PM CDT · Texas
516 people interested · 96 people going
Interested

1.9K Reactions 143 Comments

Like Comment

Heart of Texas
Sponsored
Like Page

Border Patrol agents in South Texas arrested an illegal alien from Honduras that had previously been deported and convicted of Rape Second Degree.

Thanks to Obama's and Hillary's policy, illegals come here because they wait for amnesty promised. The wrong course had been chosen by the American government, but all those politicians are too far from the border to see who actually sneaks through it illegally.

Rapists, drug dealers, human traffickers, and others. The percent of innocent poor families searching for a better life is too small to become an argument for amnesty and Texas warm welcome.

DON'T MESS WITH TX BORDER PATROL
ALWAYS GUIDED BY GOD

3.1K Reactions 89 Comments 1.2K Shares

Like Comment Share

Suggested Page

Secured Borders
Sponsored

Every man should stand for our borders! Join!

NO INVADERS ALLOWED

Secured Borders
News & Media Website
134,943 people like this.
Like Page

HATES EUROPEANS, CHRISTIANS, BUDDHISTS, JEWS, ATHEISTS, DOGS, GAYS, LESBIANS, HAM, WINE, BEER, YOUR CIVILIZATION, YOUR CULTURE, YOUR TRADITIONS, YOUR COUNTRY AND YOU.

COMPLAINS ABOUT ISLAMOPHOBIA
MERICAN FURY

Follow

Who really is the hateful one here?
#conservative #conservatives4change #takeastandnotaknee #texas #istand #trump2020 #usa #america #trump #trumptrain #trump2016 #donaldtrump #presidenttrump #draintheswamp #buildthewall #triggered #libbys #liberallogic #makeamericagreatagain #boycottnfl #2a #2ndamendment #nra #AmericaFirst #boycottnfl #democraps #maga #islam #fakenews #mypresident #american

Load more comments

how do you get through life when you are THAT paranoid? Serious question. Do you live in a war zone or something?

5,517 likes

Log in to like or comment

reality
BLACKTIVIST

The KKK has infiltrated police departments for years
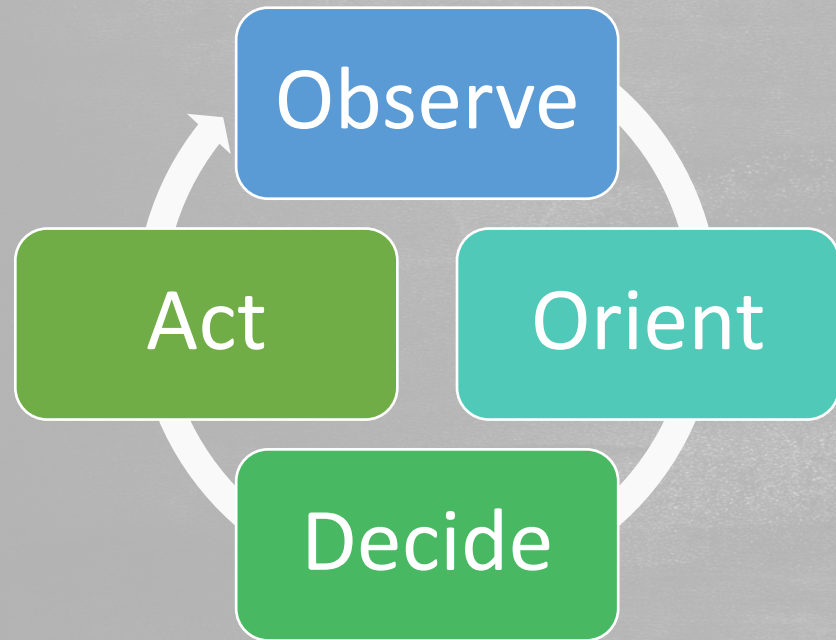
The OODA Loop

John Boyd's OODA Loop
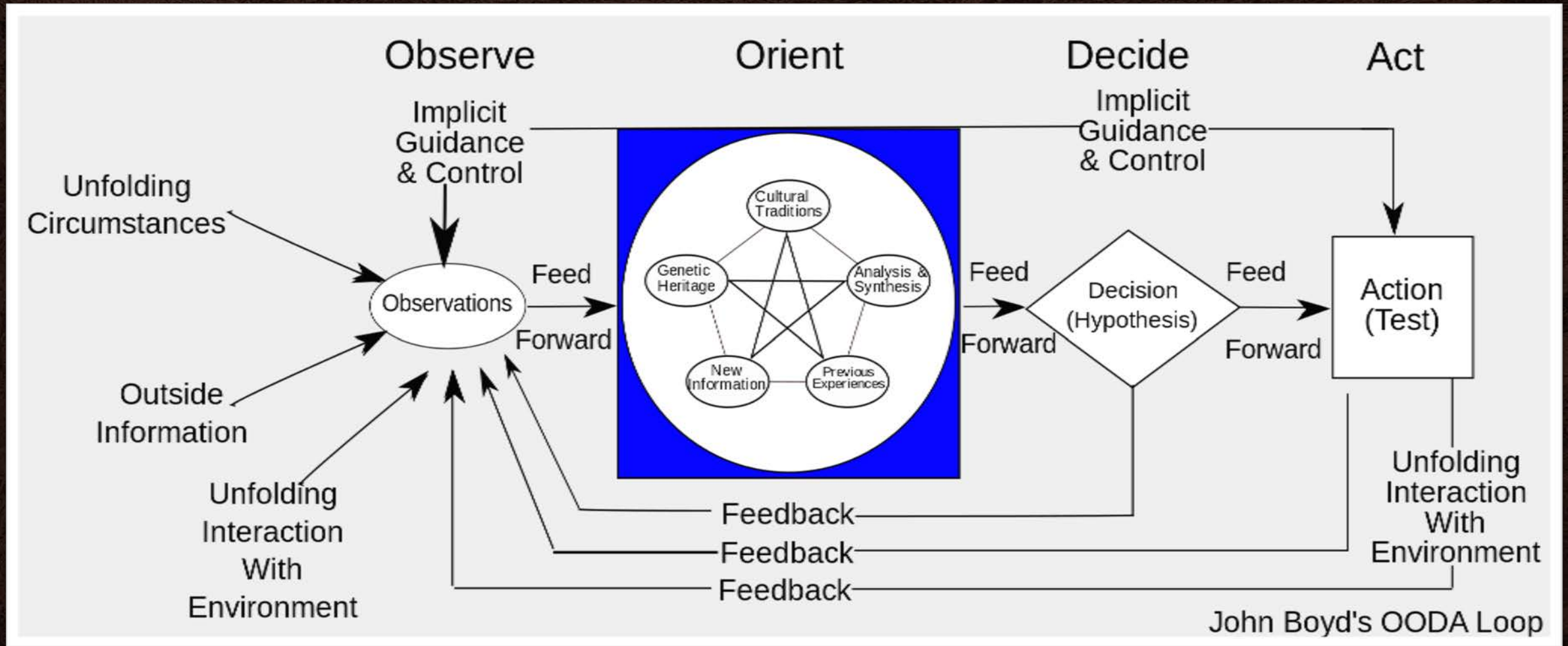
Source: https://en.wikipedia.org/wiki/OODA_loop

THE 2 SYSTEMS

READING**GRAPHICS**
ACTIONABLE INSIGHTS IN ONE PAGE

System 1 (Fast Thinking)

Continuously scans our environment.

Fast but error-prone

Works automatically & effortlessly via shortcuts, **impulses and intuition.**

System 2 (Slow Thinking)

Used for specific problems, **only if necessary**

Takes effort to analyze, reason, solve complex problems, **exercise self-control**

Slow but reliable

Graphic Source: https://readingraphics.com/book-summary-thinking-fast-and-slow/

KnowBe4
Human error. Conquered.

COGNITIVE BIAS CODEX

What Should We Remember?

- We store memories differently based on how they were experienced
- We reduce events and lists to their key elements
- We discard specifics to form generalities
- We edit and reinforce some memories after the fact
- We favor simple-looking options and complete information over complex, ambiguous options
- To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions
- To get things done, we tend to complete things we've invested time & energy in
- To stay focused, we favor the immediate, relatable thing in front of us

Need To Act Fast

- To act, we must be confident we can make an impact and feel what we do is important
- We project our current mindset and assumptions onto the past and future

Too Much Information

- We notice things already primed in memory or repeated often
- Bizarre, funny, visually-striking, or anthropomorphic things stick out more than non-bizarre/unfunny things
- We notice when something has changed
- We are drawn to details that confirm our own existing beliefs
- We notice flaws in others more easily than than we notice flaws in ourselves

- We tend to find stories and patterns even when looking at sparse data
- We fill in characteristics from stereotypes, generalities, and prior histories
- We imagine things and people we're familiar with or fond of as better

Not Enough Meaning

- We simplify probabilities and numbers to make them easier to think about
- We think we know what other people are thinking

DESIGNHACKS.CO · CATEGORIZATION BY BUSTER BENSON · ALGORITHMIC DESIGN BY JOHN MANOOGIAN III (JM3) · DATA BY WIKIPEDIA

Cognitive biases sneak in.

They underlie much of our thinking and all our assumptions.

... and they can easily be exploited!

50 Cognitive Biases in the Modern World

50 COGNITIVE BIASES TO BE AWARE OF SO YOU CAN BE THE VERY BEST VERSION OF YOU

| Memory | Social | Learning | Belief | Money | Politics |

Fundamental Attribution Error
We judge others on their personality or fundamental character, but we judge ourselves on the situation.

Sally is late to class; she's lazy. You're late to class; it was a bad morning.

Self-Serving Bias
Our failures are situational, but our successes are our responsibility.

You won that award due to hard work rather than help or luck. Meanwhile, you failed a test because you hadn't gotten enough sleep.

In-Group Favoritism
We favor people who are in our in-group as opposed to an out-group.

Francis is in your church, so you like Francis more than Sally.

Bandwagon Effect
Ideas, fads, and beliefs grow as more people adopt them.

Sally believes fidget spinners help her children. Francis does, too.

Groupthink
Due to a desire for conformity and harmony in the group, we make irrational decisions, often to minimize conflict.

Sally wants to go get ice cream. Francis wants to shop for T-shirts. You suggest getting T-shirts with pictures of ice cream on them.

Halo Effect
If you see a person as having a positive trait, that positive impression will spill over into their other traits. (This also works for negative traits.)

Moral Luck
Better moral standing happens due to a positive outcome; worse moral standing happens due to a negative outcome.

False Consensus
We believe more people agree with us than is actually the case.

Curse of Knowledge
Once we know something, we assume everyone else knows it, too.

Spotlight Effect
We overestimate how much people are paying attention to our behavior and appearance.

https://www.visualcapitalist.com/50-cognitive-biases-in-the-modern-world/

Cognitive bias is why political conversations on Facebook never go well

Note the implications across memory, social, learning, belief, money, and politics.

KnowBe4
Human error. Conquered.

Your people working from home may be
stressed and distracted.

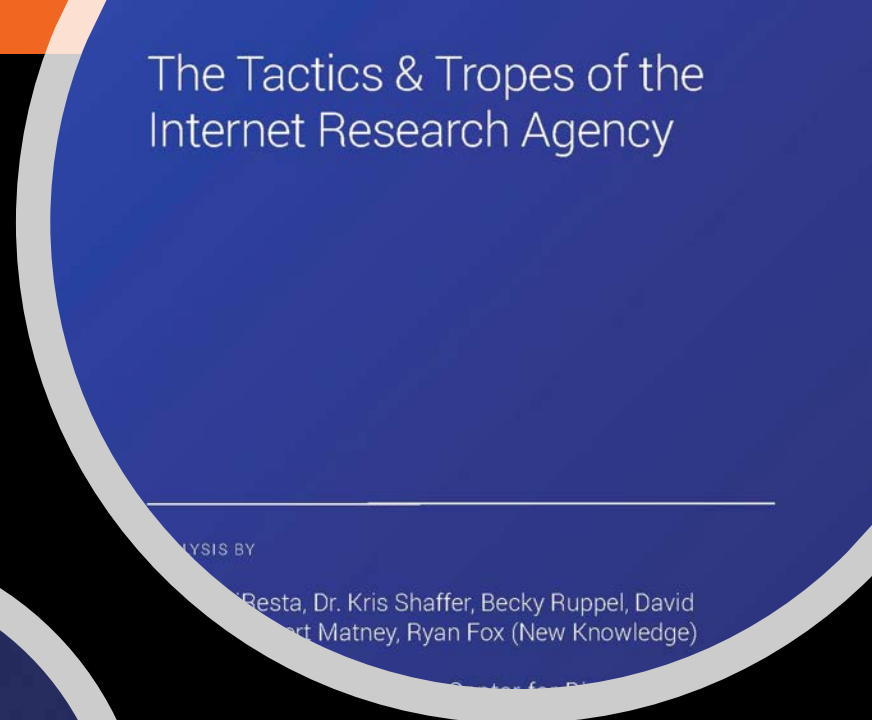…they are more vulnerable than ever
to deception and cyberattack.

Education
&
Preparation
-- are the only defense --

KnowBe4
Human error. Conquered.

40

Recommended Reading

- Manipulated: Inside the Cyberwar to Hijack Elections and Distort the Truth, by Theresa Payton

- The Reality Game: How the Next Wave of Technology Will Break the Truth, by Samuel Woolley

- The Tactics & Tropes of the Internet Research Agency

- The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation

- The Biology of Disinformation: memes, media viruses, and cultural inoculation, by the Institute for the Future

The Tactics & Tropes of the
Internet Research Agency

Resta, Dr. Kris Shaffer, Becky Ruppel, David
Matney, Ryan Fox (New Knowledge)

The Global Disinformation Order
2019 Global Inventory of Organised
Social Media Manipulation

Samantha Bradshaw · University of Oxford
Philip N. Howard · University of Oxford

MANIPULATED

Inside the Cyberwar
to Hijack Elections
and Distort the Truth

Theresa Payton

the Reality Game

How the next

KnowBe4
Human error. Conquered.

- Washington Post's Fact Checker's Guide to Manipulated Video.
  https://wapo.st/manipulatedvideo
- Fact Checking Sites
  - FactCheck.org
  - PolitiFact.com
  - www.washingtonpost.com/news/fact-checker/
  - Snopes.com
- Which Face is Real training game.
  http://www.whichfaceisreal.com/
- Indiana University Observatory on Social Media.
  https://osome.iuni.iu.edu/tools/

Your Homework: Check out the following

- **Deepfakes**:
    - **Today Show: How Deepfake Videos are Manipulating People Online.** https://www.today.com/video/how-deepfake-videos-are-manipulating-people-online-58785349692
    - **Detect AI-generated Images & Deepfakes (Part 1).** https://medium.com/@jonathan_hui/detect-ai-generated-images-deepfakes-part-1-b518ed5075f4
- **Cheap-fakes**:
    - **Real v fake: debunking the 'drunk' Nancy Pelosi footage**. https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video
- **GANs:**
    - **This Person Does Not Exist**. https://thispersondoesnotexist.com
    - **Generated Photos**. https://generated.photos
    - **Rosebud AI**. https://rosebud.ai
- **Audio Fake tools:**
    - **Descript's Lyrebird AI**. https://www.descript.com/lyrebird-ai
    - **Adobo VoCo from 2016 Keynote**. https://www.youtube.com/watch?v=I3l4XLZ59iw
    - **Google Duplex from 2018 Keynote**. https://www.youtube.com/watch?v=I2fRxfZTJ2c

Additional Recommended Resources

- **Here are the Russian Ads that Deceived Users on Facebook and Instagram.** https://techcrunch.com/gallery/here-are-15-of-the-russian-bought-ads-aimed-at-influencing-the-election/slide/1/
- **Computational Propaganda: The Book.** https://comprop.oii.ox.ac.uk/research/computational-propaganda-the-book/
- **Inside the Russian Troll Factory Where Workers Earn £980 a Week to Pump Out Putin Propaganda.** https://www.thesun.co.uk/news/6062302/inside-the-russian-troll-factory-where-workers-earn-980-a-week-to-pump-out-putin-propaganda/
- **VERIFY: People are sharing this fake picture of a Parkland survivor destroying the Constitution.** https://www.wusa9.com/article/news/local/verify/verify-people-are-sharing-this-fake-picture-of-a-parkland-survivor-destroying-the-constitution/65-531961898
- **The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public.** https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf
- **THE BIOLOGY OF DISINFORMATION: memes, media viruses, and cultural inoculation.** https://www.iftf.org/fileadmin/user_upload/images/ourwork/digintel/IFTF_biology_of_disinformation_062718.pdf
- **The Tactics & Tropes of the Internet Research Agency.** https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs

Additional Recommended Resources (Continued)

- **House of Commons Disinformation and 'fake news': InterimReport.** https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf
- **Hackers Have Already Started to Weaponize Artificial Intelligence**. https://gizmodo.com/hackers-have-already-started-to-weaponize-artificial-in-1797688425
- **The real 'fake news': how to spot misinformation and disinformation online.** https://www.cbc.ca/news/technology/fake-news-misinformation-online-1.5196865
- **Netflix documentary:** *The Great Hack*
- **HBO documentary: After Truth:** *Disinformation and the Cost of Fake News*
- **Webinar: Combating Rogue URL Tricks**. https://info.knowbe4.com/rogue-urls
- **Webinar: Learn How to Forensically Examine Emails**. https://info.knowbe4.com/phishing-forensics-krebs