Disinformation cybersecurity challenges

Date: 13 September 2023

Carbon Vector

Presented by:
Terence Fogarty
CISA | CISM | CRISC | CISSP | PCI QSA* | ISO 27001 LI | COBIT 5 Foundations
Director

Let's start off with what we know as truth

The Earth revolves around the Sun

A wet phone should be put in rice

Mount Everest is the world's highest mountain

There is zero gravity in space

Water conducts electricity

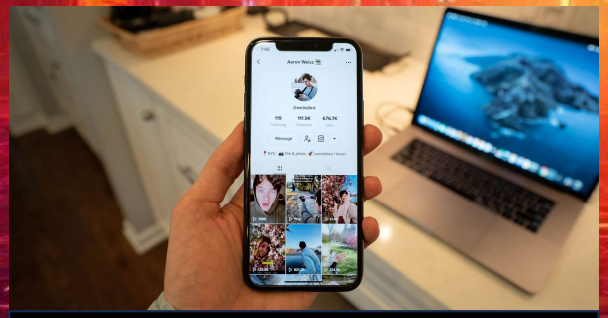
Bagpipes are Scottish

But... none of those are true.

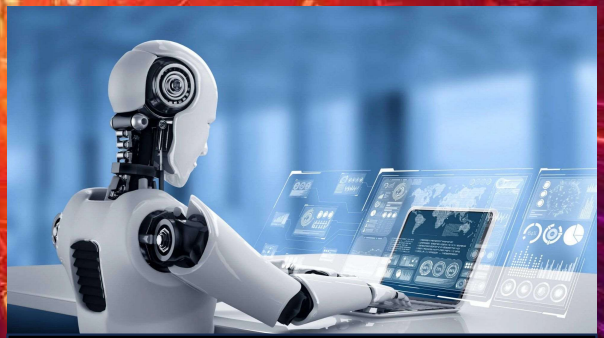
What makes us think things are true?

Parents, teachers, friends providing information

Reading it online (or used to use an encyclopedia for those over 35)



Seeing it on TV, Youtube (or Tiktok for those under 35)



Using AI (ChatGPT, Bard) tools to generate information

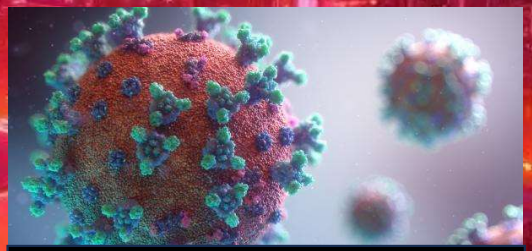
Why do we believe it?

We like drama!

Digital transformation (social media and what friends say)

Illusory Truth Effect

Confirmation bias



COVID-19 (e.g. anti-malarial drugs)

AI bias: racist, sexist, ableist and ageist

Disinformation for social engineering purposes

Missing context

- Intentionally deceptive or lacking crucial details
- Threat actors share images alongside a caption irrelevant to its content details

Deceptive editing

- Manipulates visual media, such as photos, videos, or illustrations depicting real news stories or events
- Distort reality by selectively altering essential components

Malicious transformation

- Using AI, cybercriminals can manipulate videos to fabricate realistic yet counterfeit content (“deepfakes”)
- Ransomware campaigns to amass financial profits or manipulating social consequences, e.g. election outcomes.

How is this exploited?

Attacker makes use of credentials

An attacker make use of the credentials of the information provided. This may also include information such a user credentials or personal information.

4

User enters details

User captures the details or provides the attacker with information.

3

Attack sends phishing email

Attacker crafts an email to gain information of the user, these emails are based on the trust built from the original misinformation.

2

Build trust

User clicks on the links of the misinformation and shares this data as false information. With these links being clicked on revenue is generated or used to can information of the target.

1

What can we do?

- Notifying people who shared false information
- Cybersecurity controls restricting information (and promoting cybersecurity
- Regulation of publishing information
- Determining the workings of fact checking
- Providing more context on false information (e.g. Facebook)
- Rating options for fact-checkers
- Using technology to find copies of false information
- Educate yourself and your users

Q&A

Contact details:

Terence.Fogarty@carbonvector.co.za