Week 1: introduction

Week 1 goals

This week, we’re going to cover:

- Lecture: course overview, introduction
    - Introduces students to the contents of the course, and supporting materials needed to work on disinformation data.
    - History of cognitive security
    - Working definitions of information operations, disinformation, and cognitive security
    - Disinformation examples and common myths
    - Where to find more information
- Lecture: disinformation reports, ethics
    - Example disinformation analyses.
    - Students will comment on existing disinformation analyses, and start their own research outlines.
    - How to investigate safely
    - Ethics of handling data from and about people, and groups already working on disinformation.
    - Outcome: students will be able to articulate needs and pitfalls of disinformation research, and will have started their own research outlines.
- Lecture: researcher risks
    - Potential risks to influence operations investigators, and mitigations for them.
    - Operational security
    - Mental health
    - Ethics and the golden rule (“first, do no harm”)
    - Outcome: mitigation strategies for personal risks inherent in investigating them.

Your readings are:

- No readings this week

Your assignment is:

- in-class exercise: definitions and examples of mis/disinformation, rumours, conspiracies, information operations
- Exercise: List and examine the risks in an existing influence operation.
- in-class exercise: comment on existing disinformation analyses

About this course

Class goals:

- Elaborate how information security and cognitive security interact
- Evaluate persuasive technology at different scales
- Evaluate influence operation mechanisms and tracking techniques
- Use tools to investigate account and network-level coordinated inauthentic activities
- Understand ethical behaviour around misinformation and disinformation response and research

Class prerequisite is basic python coding. Resources to catch up on this include “Python the hard way”.

Grading is on three case studies and one group project:

- Case studies (10% per study):
  - Week 2: Track a misinformation or disinformation narrative across the internet and/ or traditional media.
  - Week 6: build a disinformation landscape assessment for a country, business, or vertical
  - Week 10: gather datasets related to an existing disinformation narrative, and package them as an alert or report to be sent to a disinformation response group.
- Group project (20%):
  - Apply cognitive security techniques to an incident, country, business, or community.
  - Identify a problem, use the lens of one of the frameworks we studied in class to address it, and explore/analyze *theoretically* how to successfully measure and counter it.

Defining Cognitive Security

The definition of Cognitive Security used in this class is: “Cognitive security is the application of information security principles, practices, and tools to misinformation, disinformation, and influence operations. It takes a socio-technical lens to high-volume, high-velocity, and highvariety forms of “something is wrong on the internet”. Cognitive security can be seen as a holistic view of disinformation from a security practitioner’s perspective”.

The term Cognitive Security comes from two different places:

- MLsec: “Cognitive Security is the application of artificial intelligence technologies, modeled on human thought processes, to detect security threats.” — XTN. This is the MLsec definition, of machine learning in information security — in attack, defence, and attacking the machine learning systems themselves. This is adversarial AI, and Andrade2019 is a good summary of this field.
- Social engineering: “Cognitive Security (COGSEC) refers to practices, methodologies, and efforts made to defend against social engineering attempts–intentional and unintentional manipulations of and disruptions to cognition and sensemaking” — cogsec.org. This version of the term, coined by Rand Waltzman, is the social engineering at scale definition, about manipulating individual beliefs, sense of belonging etc, and manipulation of human communities. This could be seen as adversarial cognition, and Waltzman2017 and the COGSEC.org website created after his testimony are good summaries of it.

These definitions aren’t as incompatible as they look: they’re both based on adversarial activities, and defence against the manipulation of information, knowledge, and belief. But neither of them quite capture what’s going on today, where we’re seeing both humans and algorithms being manipulated to changes the fates of individuals, communities, organisations, and countries, although as I write this, I could see that the second definition could include algorithms if we allow cognition and sensemaking to cover algorithms too.

Both of these definitions are from the point of view of defence — something that was a strong driver of our (the CredCo MisinfosecWG) own adoption of a term that included “security”, but feels less appropriate when we’re modelling influence in information ecosystems, and what we’re looking at seems more and more to resemble massive multiplayer games, where each individual, community, organisation, country etc has its own goals, and may see even the most aggressive influence actions as part of defending its own realm. MLsec is helpful here, with its separation into study of attacks using ML (machine learning algorithms), defence using ML, and attacks on the ML processes themselves (Bruce Schneier’s paper on common knowledge attacks against democracy fits the latter part). It’s useful to be aware that your cognitive security defence moves might be viewed as someone else’s attack.

There are also two definitions of social engineering:

- Centralised planning: “the use of centralized planning in an attempt to manage social change and regulate the future development and behavior of a society.” — basically mass manipulation
- Individual deception: “the use of deception to manipulate individuals into divulging confidential or personal information that may be used for fraudulent purposes.” — basically phishing etc

Both of these are compatible with the definitions of cognitive security above. I think the definitions are vague enough to also cover something else that gets lost sometimes: that the entity being manipulated isn’t just knowledge (“truth” etc), but also include manipulation of group cohesions (“belonging”) and emotions (“feels”), both of which can be changed with information that’s completely true.

The centralised planning definition is interesting as we shift from responding to disinformation incidents one-by-one, to discussing how to improve our information environments (e.g. by making verified information easier to find online), and hopefully creating resilience at all levels rather than mandating it from above. In spaces where many entities are competing for attention and influence, viewpoints matter, autonomy and individuals matter, and resilience and vulnerability are most likely to stem first at the individual and community level.

The Roots of Cognitive Security

We will be dealing with misinformation, disinformation malinformation, rumours, and conspiracies.   These are all part of the CogSec threat landscape, and have all been ‘found’, highlighted, studied, and countered/ mitigated by different communities.

- The media community focussed on misinformation and disinformation. Clare Wardle’s “types of information disorder” diagram showed them as a venn diagram of falseness and intent to harm, where misinformation was falseness without intent, malinformation was intent without falseness, and disinformation was falseness and intent. This was very content-based, because a lot of the early focus was on not polluting media articles that had started to use User-Generated Content (after Web2.0, anyone could post anywhere, and using this content was an easy fix for media funding woes).
- The military community focussed on psyops (renamed MISO: military information support operations). GAO’s diagram of the US Department of Defense showed this as part of information operations, alongside military deception, cyberspace operations, electromagnetic warfare, special technical operations, and operations security.
- Targetted communities (technical women, Black Americans etc) focussed on surviving GamerGate-style personal attacks. They built backchannels and coping strategies long before many other communities noticed there was a problem. Shireen Mitchell’s Stop Online Violence Against Women group is a good example of these.
- The information security community focussed on social engineering at scale. Our own diagram of information security being split into physical, cyber, and cognitive security is a nod to the many foundational information security texts that included human cognition, in various forms, from the start.
- There are other communities in the cognitive security space, but these four drove a lot of early work.

Things we've borrowed from Information Security

One way of looking at Cognitive Security is as a parallel effort to cyber security, but with brains, beliefs, and communities substituted for computers, data, and networks. Although there are differences between these domains that are pointed out throughout the course, this analogy has served us well in finding cybersecurity ideas that might help with things like disinformation defence. Despite these, we're still dealing with two domains that are carried on the Internet, and usually result in actions.

- One early idea borrows from the CIA triad of confidentiality (only the people/systems that are supposed to have the information do so), integrity (the information has not been tampered with), and availability (people can use the system as intended). Danny Rogers first pointed out that disinformation is an integrity problem, where beliefs, belonging etc have been tampered with.
- Another useful idea is adapting the ATT&CK framework to model and manage disinformation creator and responder behaviours (aka TTPs, or Tactics, Techniques, and Procedures). This became the AMITT set of disinformation behaviour models.
- The STIX model of information security actors, behaviours, content, tools, indicators, vulnerabilities, and infrastructure also adapted easily to disinformation use, with only two minor changes (adding a narrative object that mirrored the use of malware objects, and an incident object that behaved similarly to an intrusion set).
- Viewing disinformation as a risk management problem has been extremely useful, allowing us to do similar analyses to those seen in other parts of information security risk management: quantifying and assessing risks (how bad, how big, who to), and components including attack surfaces, vulnerabilities, potential losses and outcomes. This allows for risk assessment, reduction, and remediation, but more importantly, in an era when misinformation, disinformation, and rumours are everywhere, helps us answer the question of where to put detection, mitigation, and response resources, where resources include people, technologies, time, attention, and connections. As far as I know there isn't a disinformation version of the FAIR risk management framework yet, but it's not a big adaptation (a few category changes), so that's probably only a matter of time too.
- Another thing borrowed is the idea of tiered security operations centers: their structure, activities, resources, and principal objects. These have proved themselves useful in the past year.

What we're dealing with

Risk and Safety

Disinformation Reports

Week 1 References

Readings

- No readings

Background

- Python the hard way
- Clare Wardle, Hossein Derakhshan, "information disorder: toward an interdisciplinary framework for research and policy making", Council of Europe report, 2017
- Joseph Kirschbaum, "information environment. DoD operations need enhanced leadership and integration of capabilities", GAO Testimony, April 30 2021 (also see https://www.gao.gov/products/gao-21-525t)
- https://xtn-lab.com/what-is-cognitive-security/
- cogsec.org
- https://languages.oup.com/google-dictionary-en/
- Parkerian Hexad: https://www.staffhosteurope.com/blog/2019/03/cybersecurity-and-the-parkerian-hexad and https://www.sciencedirect.com/topics/computer-science/parkerian-hexad
- Danny Rogers, Black Hat / Forbes
- OII, Computational propaganda report
- https://euvsdisinfo.eu/disinformation-cases/ - Russia disinfo on EU
- https://medium.com/dfrlab - world disinfo
- https://comprop.oii.ox.ac.uk/ - nationstate actors
  - Specifically The Global Disinformation Order and https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/Case-Studies-Collated-NOV-2019-1.pdf
- https://www.newsguardtech.com/covid-19-resources/ - c19 domains for several countries
- Graphika, "IRA in Ghana: double deceit", 2020
- CISA, "the war on pineapple"
- https://dai-global-digital.com/cyber-harm.html
- EFF's Surveillance Self-Defense guide https://ssd.eff.org/en/module/your-security-plan
- Security Guidelines for Congressional Campaigns https://techsolidarity.org/resources/congressional_howto.html
- https://www.vice.com/en_us/article/a37p94/what-is-threat-modeling
- Botsentinel: themes "trollbots" are promoting
- Hamilton68 - public version is feeds from official Russian sites (embassies, RT etc), not trolls.

- Ryerson University covid19 misinformation portal
    - Botswatch dashboard Botswatch dashboard
- Indiana University OSOME Decahose
- Facebook Datafeed
- Uni Arkansas COSMOS Covid19 list
- Wikipedia list of Covid19 rumours
- WHO Covid19 myths list - narratives
- Ryerson Claimwatch dashboard
- CMU IDEAS Center list of Covid19 disinformation narratives
- Indiana Hoaxy (twitter, articles)
- EuVsDisinfo database https://euvsdisinfo.eu/disinformation-cases/
- Atlantic Council DFRLab https://medium.com/dfrlab
- Graphika https://graphika.com/reports
- Facebook https://about.fb.com/news/2021/08/july-2021-coordinated-inauthentic-behavior-report/
- FireEye https://www.fireeye.com/blog/threat-research/2020/07/ghostwriter-influence-campaign.html
- Many many factchecking orgs https://datastudio.google.com/reporting/a8491164-6aa8-45d0-b609-c70339689127/page/ierzB
- https://twitter.com/brechtcastel/status/1431612326759829513?s=19
- https://twitter.com/conspirator0
- https://www.vice.com/en/article/93yvmv/qanon-ghostezra-is-robert-randall-smart
- https://www.atlanticcouncil.org/in-depth-research-reports/the-long-fuse-eip-report-read/
- https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/how-to-report-misinformation-online