KNOWLEDGE MANAGEMENT IN A POST-TRUTH WORLD:

How can we improve the health of our knowledge ecosystem?

DR. KIMIZ DALKIR, PH.D.
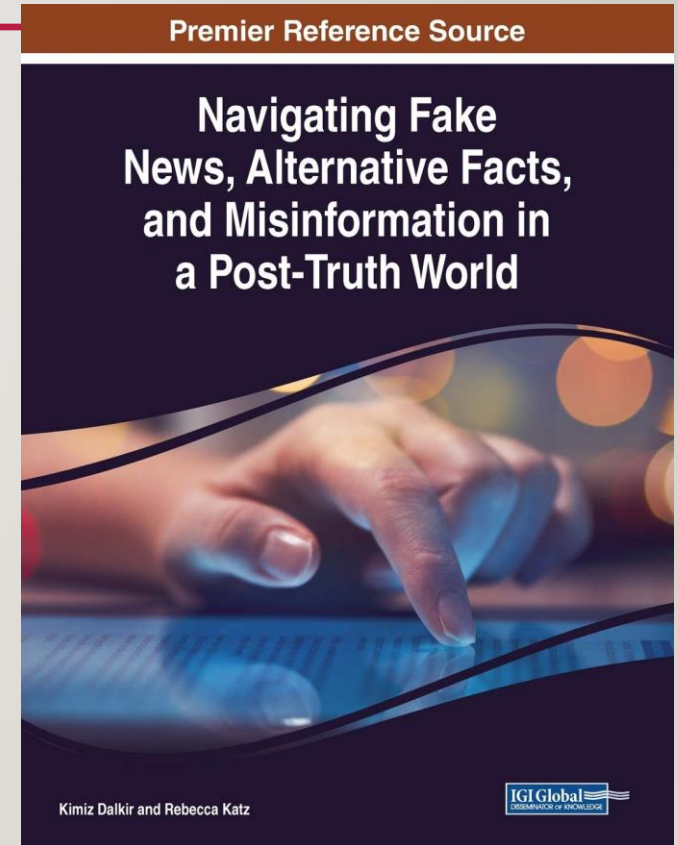
4TH ANNUAL DIGITAL TRANSFORMATION IN GOVERNMENT CONFERENCE
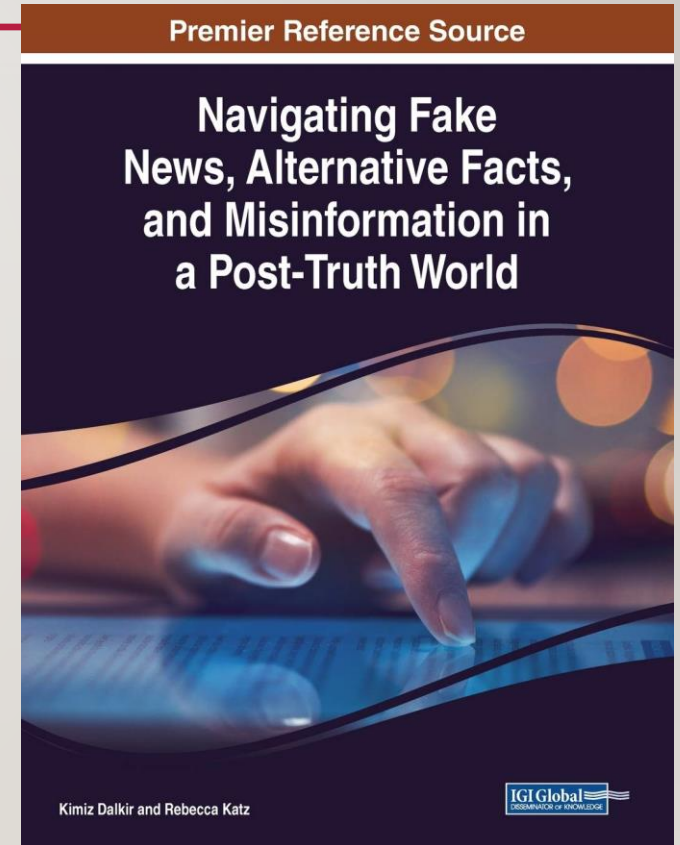
13-16 JUNE 2023, OTTAWA

OVERVIEW

- What is fake news?
  - Brief history of post-truth – is it really new?
- Role of information/media literacy
- Role of organizational information policy and legislation
- How can KM (and AI) help?
- Recommendations

Premier Reference Source

Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World

Kimiz Dalkir and Rebecca Katz

IGI Global
DISSEMINATOR OF KNOWLEDGE

OVERVIEW

- ## What is fake news?
  - ### Brief history of post-truth – is it really new?
- ## Role of information/media literacy
- ## Role of organizational information policy and legislation
- ## How can KM (and AI) help?
- ## Recommendations

Premier Reference Source

Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World

Kimiz Dalkir and Rebecca Katz

IGI Global
DISSEMINATOR OF KNOWLEDGE

WHAT IS POST-TRUTH?

- In a post-truth world, we are deliberately looking for facts that support our beliefs, opinions & conclusions…

- Post-truth is the "public burial of objective facts by an avalanche of media appeals to emotion and personal belief" *(Oxford Dictionary)*
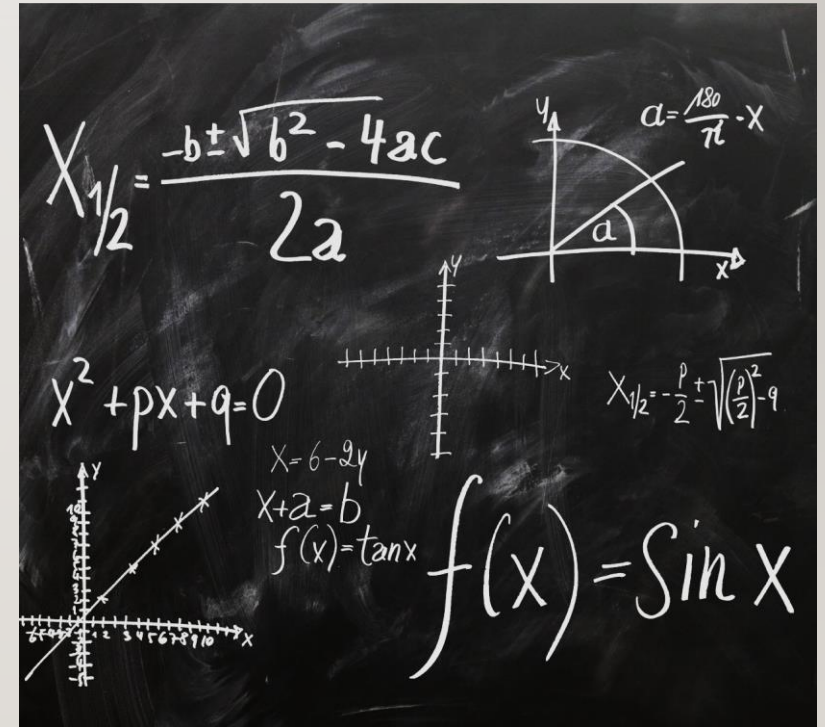
We are prone to confirmation bias as we believe and repeat statements that support our views (Enfield)

(BRIEF) HISTORY OF FAKE NEWS

- This is not a new phenomenon but….the advent of the Internet and social media have definitely increased the geographic reach (it is worldwide) and the speed (almost instantaneous) with which we share fake news … as well as the use of AI to generate fake news

A timeline of post-truth

[< 2000]	[2000+]	[2016-2017]	[2018]	[2018+]
One-to-many communications:	Networked systems	Peak?; US elections	First legislation with fines (Germany)	First fact checkers

gossip, rumours e.g. George Orwell's War of the Worlds

Online fake news

COVID 19

Did Nero really fiddle while Rome burned?

https://www.bbc.co.uk/bitesize/articles/zwcgn9q

SO WHAT IS DIFFERENT TODAY?

- Information Overload
  - So many channels pushing information at us in real time
  - "we live in a world where there is more and more information and less and less meaning" (Baudrillard, 1994, p. 79)

- Information Avoidance" ready to ignore facts, not accept claims

- Lack of time to verify *everything* coping with information overload

Info Overload + Info Avoidance – Time = Post truth

CROWD-SOURCED CONTENT

- Most people read news through their social networks not newspapers, newsfeeds or television news (over 66% from Meta in US).

GroupThink – fastest speed of sharing is between close networks of trusted colleagues, family and friends – the more it is repeated the more it is believed ("Repetition theory (Hasher et al, 1977)

Trust + Share Speed + Repetition = Post truth

FILTER BUBBLES (PARISER, 2011)

- Filter bubbles are applied to all contemporary mediated content

- Everyone lives in their own bubble of content created by personalization algorithms of the social media services (artificial intelligence used to recognize your patterns)

- In this way, you only receive content that you would react to favorably or find interesting

- Same algorithms used in database marketing to sell you things you are likely to buy

- We create our own social media bubbles and choose to live in them - - but the **scale** today is exponentially greater and the **speed** of sharing takes only one click or "like"

Value is measured by reach (#likes) not validity

AI-GENERATED FAKE NEWS

- Smart tools that allow computers to pose as humans

- Text-generating AI that produces synthetic text ("readfakes")

- AI-generated photos and videos ("deepfakes") are getting much harder to detect – we can make anyone say anything including former Presidents

- AI s/w was much more successful at getting users to click on malicious links than human could

AIs can generate fake reports that are convincing enough to trick cybersecurity experts (not just the general public)

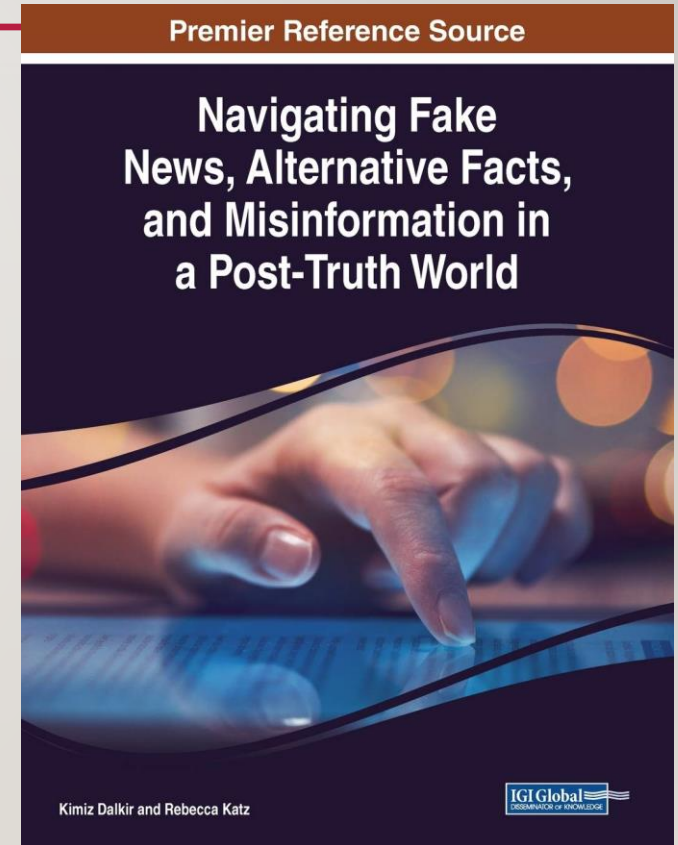https://www.youtube.com/watch?v=AmUC4m6w1wo&ab_channel=BBCNews

CHATGPT

- Can not only generate large volumes of content very quickly but a amount of that content will not be valid

- Generative AI will also soon be used for your online search queries
    - Again – not necessarily validated so will direct traffic to misinformation
    - Also will tend to just give you an answer instead of directed you to a site
    - Best at creating simple short summaries that are devoid of nuance (let alone critical analysis)
    - Google's new generative AI program (not yet released in Canada) provides shortsummaries of blogs and articles for your search (https://www.ctvnews.ca/sci-tech/toronto-blogger-warns-of-seismic-shift-in-seo-amid-ai-advancements-1.6421036)

OVERVIEW

- What is fake news?
  - Brief history of post-truth – is it really new?
- **Role of information/media literacy**
- Role of organizational information policy and legislation
- How can KM (and AI) help?
- Recommendations

Premier Reference Source

Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World

Kimiz Dalkir and Rebecca Katz

IGI Global
DISSEMINATOR OF KNOWLEDGE

WHAT MAKES PEOPLE VULNERABLE TO MISINFORMATION?

- Perception that the content is benign

- Living in our filter bubbles

- Lacking information (or "digital") literacy skills
  - Taking the time and making the effort to fact check

- Laziness

- Not enough times

- We are not as rational in our thinking as we would like to believe…

Image source: https://news.stanford.edu/2021/10/25/foil-fake-news-focus-infectiousness/

DIGITAL CITIZENSHIP RESEARCH PROJECT (HERITAGE CANADA)

- Looked at ways in which we could inoculate marginalized youth against fake news

- Increased media literacy using mobile phone app that used short games

- Marginalized youths in particular tend to avoid traditional media outlets and have greater information avoidance behaviours.

- Found playing the fake news game reduced the perceived reliability and persuasiveness of fake news
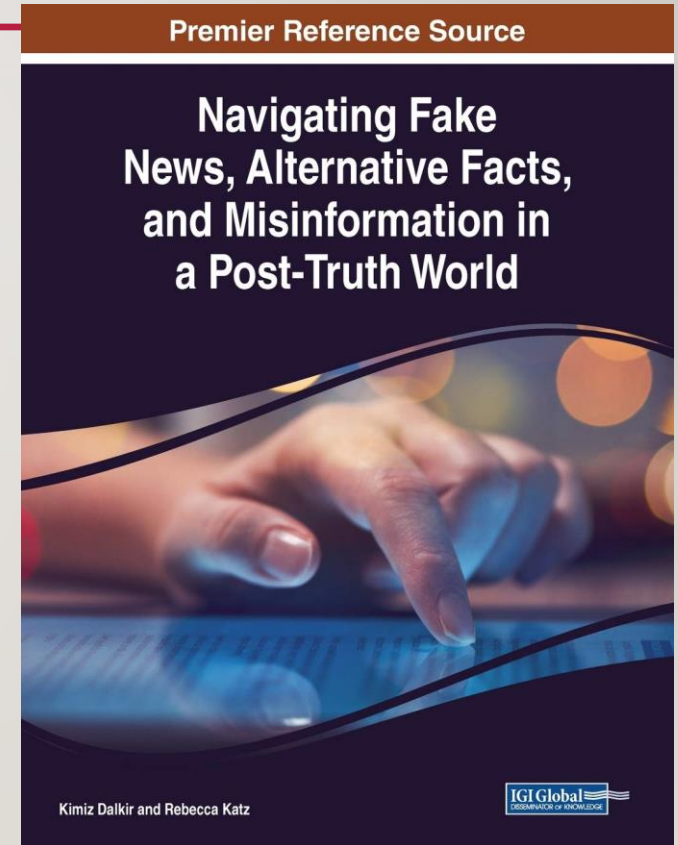
- Active inoculation (participation) was more successful than passive reading

Early media education & educational games may therefore one way to inoculate people against the risks of disinformation

WHAT CAN HELP?

- Beyond information literacy – metaliteracy,
  - Digital literacy
  - News literacy
  - Multiple literacies
  - Learn how to assess the credibility of sources, triangulating (gathering data from multiple sources and see if they agree or refute), critical reasoning skills
  - Learn how to spot fake websites, altered images, misinformation, fraudulent requests by looking for clues such as inconsistencies in lighting…

OVERVIEW

- What is fake news?
  - Brief history of post-truth – is it really new?
- Role of information/media literacy
- **Role of organizational information policy and legislation**
- Role of technology
 - Social media & Artificial intelligence (AI)
- Recommendations

Premier Reference Source

Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World

Kimiz Dalkir and Rebecca Katz

IGI Global
DISSEMINATOR of KNOWLEDGE

WHAT ABOUT INFORMATION POLICIES?

- Social media rules on misinformation vary widely & are often confusing, unclear, or contradictory

- Hard to know what to expect on each platform, and to choose where to go for important information.

- Consumer Reports analyzed misinformation policies from the country's biggest social media platforms
  - focused on the most dangerous types of falsehoods, including misinformation on the coronavirus and on how to vote when the consequences are serious

Most of the platforms don't have a blanket rule against posting false material, but they do ban certain kinds of misinformation

https://www.consumerreports.org/social-media/social-media-misinformation-policies/
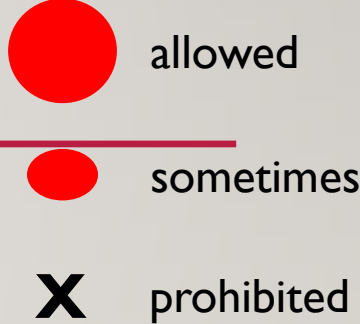
TIKTOK COMMUNITY GUIDELINES ON MISINFORMATION
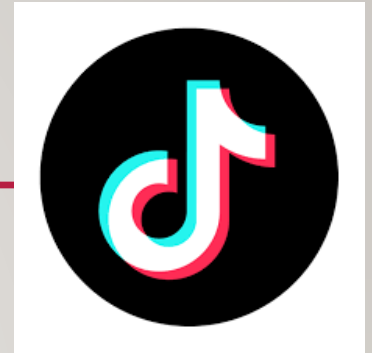
Do not post, upload, stream, or share:

- Misinformation that incites hate or prejudice

- Misinformation related to emergencies that induces panic

- Medical misinformation that can cause harm to an individual's physical health

- Content that misleads community members about elections or other civic processes

- Conspiratorial content including content that attacks a specific person or a protected group, includes a violent call to action, or denies a violent or tragic event occurred

- Digital Forgeries (Synthetic Media or Manipulated Media) that mislead users by distorting the truth of events and cause significant harm to the subject of the video, other persons, or society

Do not engage in coordinated inauthentic behavior such as the use of multiple accounts to exert influence and sway public opinion while misleading individuals, our community, or our systems about the account's identity, location, relationships, popularity, or purpose

Which platforms allow false information

	Politics/ Social	Health/ Coronavirus	Voting/ Census	Manipulated Media
Facebook/ Instagram	● (allowed)	● (sometimes)	✕	✕
YouTube	● (allowed)	● (sometimes)	✕	● (sometimes)
Twitter	● (allowed)	● (sometimes)	✕	● (sometimes)
Pinterest	● (sometimes)	✕	✕	✕
Reddit	● (allowed)	● (sometimes)	● (allowed)	✕
Snapchat	● (sometimes)	✕	✕	✕
WhatsApp	● (allowed)	● (allowed)	● (allowed)	● (allowed)
TikTok	● (sometimes)	✕	✕	✕

● allowed
● sometimes
✕ prohibited

RECENT CHANGES

- YouTube stops deleting false 2020 election claims(https://blog.youtube/inside-youtube/us-election-misinformation-update-2023/)
    - will remove references to election fraud but continue to remove misinformation on how to vote in 2024
- Meta starts testing paid verification for Canadian Facebook, Instagram creators (https://www.ctvnews.ca/sci-tech/meta-starts-testing-paid-verification-for-canadian-facebook-instagram-creators-1.6421340).
    - Pay to be verified ("Meta-verified" blue check mark)
    - approx. $20/month using govt id
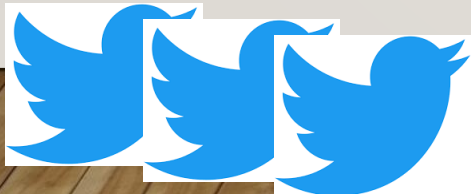
Twitter?

WHAT ABOUT LEGISLATION?

- Countries need legislation on fake news that does not infringe upon human rights and free speech BUT lack of clear definition of fake news and the requirement to prove malicious intent makes this difficult

  - Potential for abuse is high especially if people can be prosecuted, fined or jailed for spreading fake news

- Some other approaches include:

  - Media literacy campaigns, government task forces; websites, online portals, government fact-checking sites where you can report fake news and have it validated, database of fake news stories distributed on social media and apps that let people report misinformation and apps that use AI to detect fake news

https://www.poynter.org/ifcn/anti-misinformation-actions/

GERMANY (2018) – THE PIONEER (JUNE 2017)

- "Netzwerkdurchsetzungsgesetz" (NetzDG) Law forces online platforms to remove "obviously illegal" posts within 24 hours or risk fines of up to €50 million

- Aimed at social networks with more than 2 million members — such as Meta, YouTube and Twitter (enforcement has been rocky)

- Problems e.g., a satirical magazine called Titanic published a piece that was banned from Twitter, and even the minister of justice — who helped author the NetzDG — had his tweets censored

- Following criticism that too much content was being blocked

INTERNET CENSORED

FRANCE

- The Election Misinformation law (2018): rules on media to prevent "Inexact allegations or imputations, or news that falsely report facts, with the aim of changing the sincerity of a vote" in the three months preceding any vote

- Judge has power to remove fake content spread via social media and block the sites that publish it and must act within 48 hrs of notification

- The Higher Audiovisual Council (CSA), the broadcasting regulator, can revoke the broadcast rights of French media outlets found to work "under the control or influence of a foreign state to disseminate misinformation
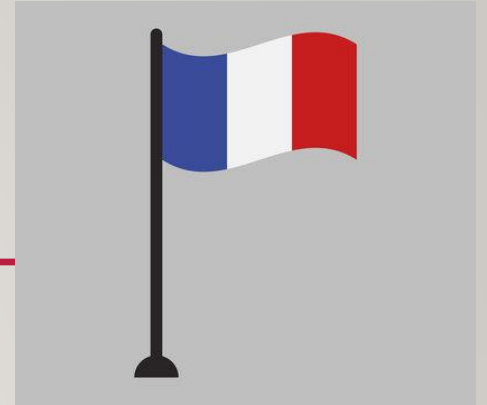
Text later amended to target the "manipulation of information" instead of "fake news" in order to protect satire from being penalized under this law ☺

SINGAPORE: THE PROTECTION FROM ONLINE FALSEHOODS AND MANIPULATION ACT

- Latest to introduce a law but among the most comprehensive in the world

- Illegal to spread "false statements of fact" that compromise security, public tranquility, public safety and the country's relations with other nations

- $37,000 or five years in prison if you share fake news

- $74,000 and a potential 10-year jail term if fake news shared via fake online account or a bot

- $740,000 and potential 10 years jail term if social media platforms play a role in spreading misinformation

- Criticized for limiting free speech & got pushback from Meta

U.S. MEDIA LITERACY LEGISLATION

- California state government passed a law to promote media literacy in public schools
  - Department of Education must list instructional materials and resources on how to evaluate trustworthy media
  - Law was inspired by a Stanford University student who found that most students can't distinguish between sponsored content and news stories
- Massachusetts state government passed a similar bill that mandates civic education with an emphasis on media literacy
- Most if not all date from 2018 following highly publicized fake news intended to influence elections

CANADIAN GOVERNMENT APPROACHES

- Bill C-76 tech companies must be more transparent about their anti-disinformation and advertising policies

- Created a "Critical Election Incident Public Protocol" that will monitor and notify other agencies and the public about disinformation attempts.

- **Announced it was giving $7 million to projects aimed at increasing public awareness of misinformation online**

- Launched digital charter: "The Government of Canada will defend freedom of expression and protect against online threats and disinformation designed to undermine the integrity of elections and democratic institutions."

- PM implied "meaningful penalties" for tech companies that don't comply, but the charter does not outline how fines would work, nor does it define fake news.
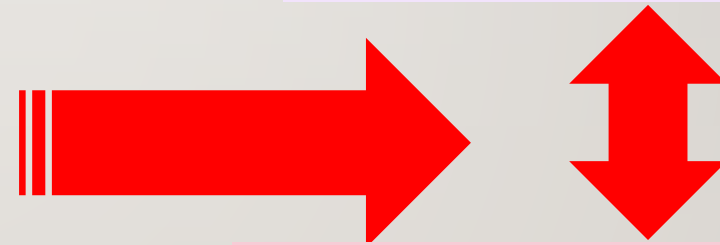
OVERVIEW

- What is fake news?
  - Brief history of post-truth – is it really new?
- Role of information/media literacy
- Role of organizational information policy and legislation
- **How can KM (and AI) help?**
- Recommendations

Premier Reference Source

Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World

Kimiz Dalkir and Rebecca Katz

IGI Global
DISSEMINATOR OF KNOWLEDGE

HOW CAN KM HELP?

.

Machine AI

Human trust

To manage information is to filter it - - better

KM processes can be used at the individual, group/community and organizational levels to filter, validate, preserve and act upon valid content.

McElroy (2003)

Every item in a knowledge management system starts off as a "knowledge claim" - - needs to be validated, tagged and recommended before it can be used
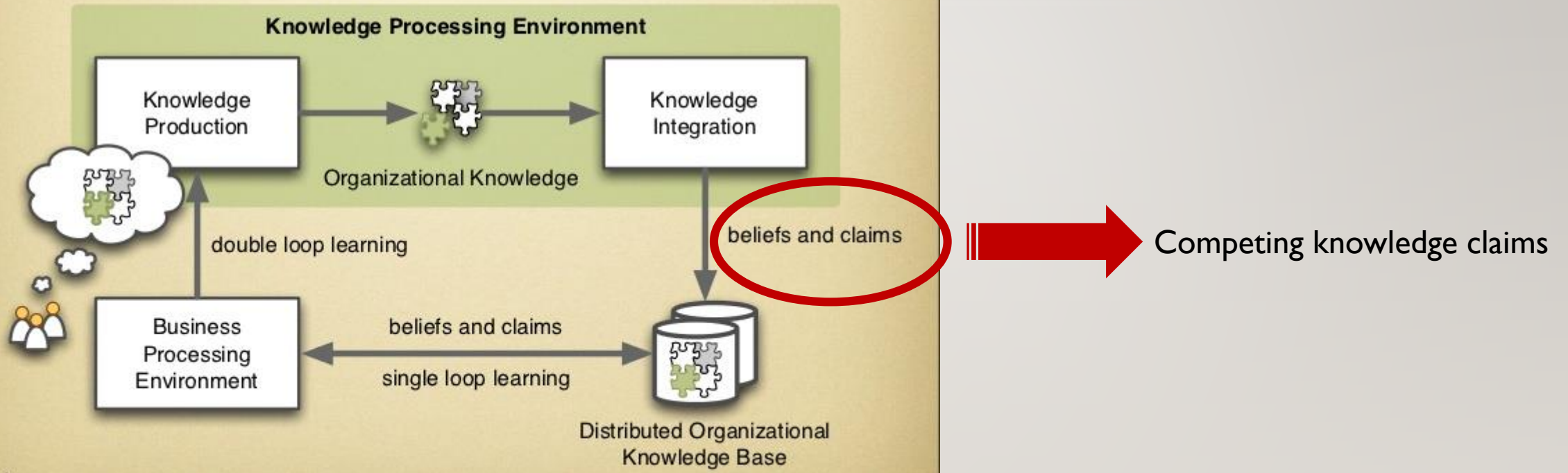
McElroy Knowledge Management Lifecycle/Process Model *(Dalkir, 2023)*

SCOTTISH JUSTICE SYSTEM: 3 VERDICTS

- Knowledge claims are either accepted, rejected or labelled "not proven." The community of trusted users provide feedback via the KM system when they use the knowledge.

"Not Proven"

# AI IS REALLY GOOD AT CREATING FAKE CONTENT – HOW CAN IT HELP DETECT (AND EVEN BETTER) PREVENT THIS?

- we can reverse engineer good solutions from great forgeries

- Other approaches:
  - Fake News uses lots of emotional words like "hate" "shocking" "secret" …. that we can identify (called "sentiment analysis") Also tend to have more photos, videos than text
  - Distinctive dissemination pattern of fake news over social media

AI is really good at creating and disseminating fakes news BUT AI can also help " fight fire with fire"
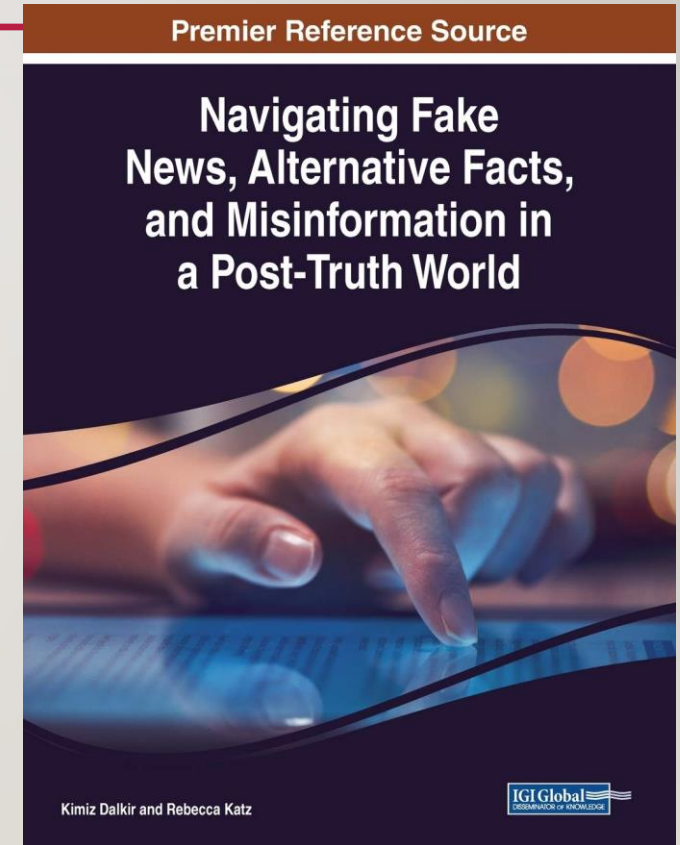
AI TOOLS TO DETECT FAKE CONTENT

- Can analyze images, videos, text and produce a recommendation such as:
  - 97% probability that this is video, photo, news item, etc. is fake
  - Pattern recognition to compare to authentic content, detect discrepancies invisible to the human eye (lighting differences, distance from camera changes, they don't blink!)
  - REMEMBER: this is what AI does best, and does better than us: looking at enormous amounts of data to make sense of it

Fake Image Detection Sites | Fake or Real ?
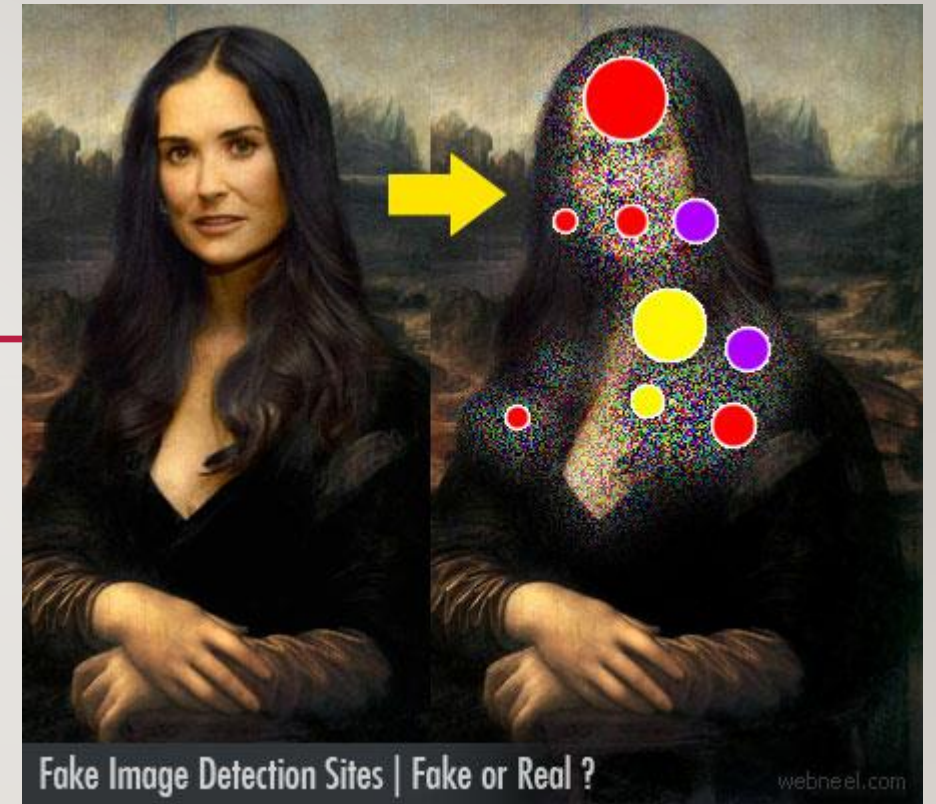
Fake video detection

OVERVIEW

- What is fake news?
  - Brief history of post-truth – is it really new?
- Role of information/media literacy
- Role of organizational information policy and legislation
- How can KM (and AI) help?
- **Recommendations**

Premier Reference Source

Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World

IGI Global
DISSEMINATOR OF KNOWLEDGE

Kimiz Dalkir and Rebecca Katz

THE BEST WAY FORWARD?

- Don't rely on AI alone nor human reasoning alone
  - Each can make mistakes
  - **Together they make a stronger team**
- New research is using the analogy of nutritional labels
  - Don't take away human agency – we are ultimately responsible in deeming content as true or false (or "unproven")
  - What we are working on now: AI to provide enough information on credibility, indicate any red flags in the content so that we can **make more informed decisions**

Maple syrup nutritional label

THE BEST WAY FORWARD IS HYBRID

- AI can help provide enough information on source credibility and indicate any red flags in the content   (Example: red flag warnings for unsafe websites)
  - BUT: do not remove human agency – we are ultimately responsible in deeming content as true or false
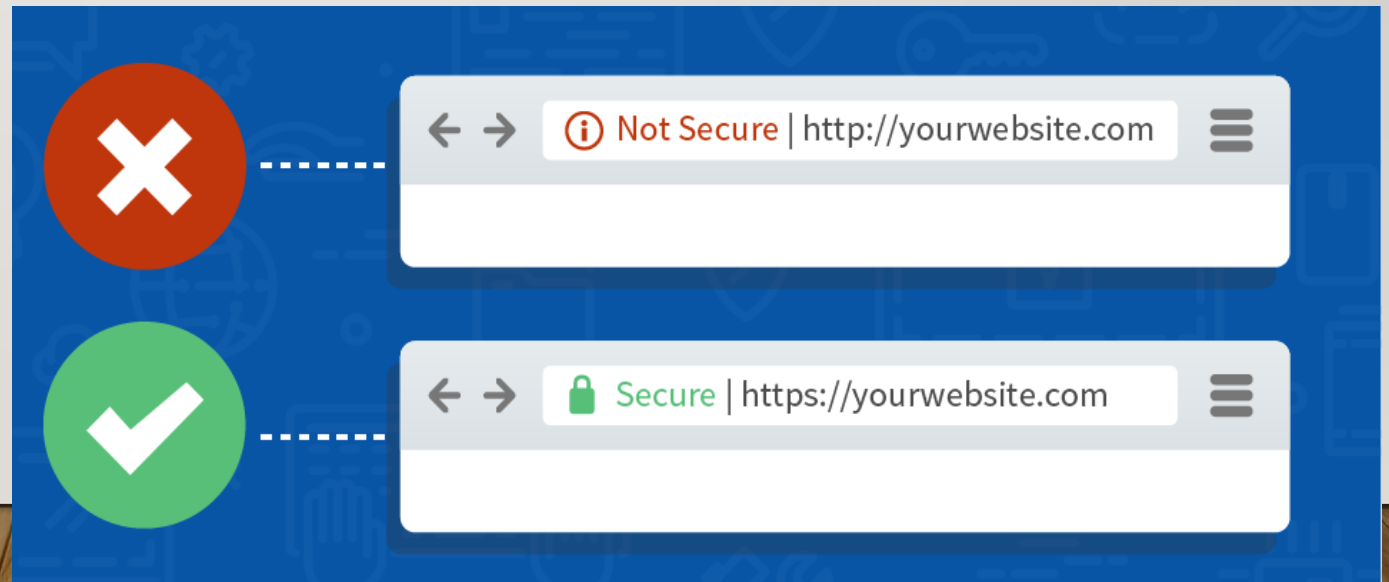
KEY TAKEAWAYS:

- Increase literacy of content consumers

- KM & AI can provide a good starting point (e.g. nutritional labels)

- KM & AI can detect and flag suspicious content

- KM & AI can detect signature sharing patterns specific to fake news

- Have a "healthy dose" of skepticism if content intrigues you

- Decide about knowledge claims yourself

- People can then validate & decide if they can trust their community (filter bubbles)

- Think before you share

SOME RESOURCES

- Baudrillard, J. (1994). Simulacra and Simulation. Translated by Sheila Faria Glaser. Michigan: Univerity of Michigan.

- Dalkir, K. (2023). *Knowledge Management in Theory and Practice*. 3^{4th} Edition. Boston, MA: MIT Press

- Enfield, N., (2017). We're in a post-truth world with eroding trust and accountability. It can't end well. *The Guardian*. Available at https://www.theguardian.com.

- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82-89

- Fuhr et al, 2020]. Fuhr, Norbert, et al. "An information nutritional label for online documents." *ACM SIGIR Forum*,51(3). New York, NY: ACM.

- Hasher, Lynn; Goldstein, David; Toppino, Thomas (1977). "Frequency and the conference of referential validity" *Journal of Verbal Learning and Verbal Behavior,* 16(1): 107–112.

- McElroy, M. W., & Mcelroy, L. (2003). *The new knowledge management: Complexity, learning, and sustainable innovation*. Routledge.

- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

- Wendling, M. (2018). The (almost) complete history of "fake news." *BBC Trending*. Https://www.bbc.com/news/blogs-trending-42724320.

Kimiz.Dalkir@mcgill.ca