Teaching About Deepfake in a Digital History Course

Dr. Noa Yaari

noa.yaari@utoronto.ca

CSDH/SCHN | Congress | Montreal, 21 June 2024

Travel grants by CSDH/SCHN and U of T's New College and Critical Digital Humanities Initiative

Digital History course - class Typos, Errors, and Deepfake - focusing on the latter

- Novel digital technologies make it increasingly difficult to distinguish between real and fake media.
- One of them is deepfakes which are hyper-realistic videos that apply AI to depict someone say and do things that never happened.
- Coupled with the reach and speed of social media, convincing deepfakes can quickly reach millions of people and have negative impacts on our society (Westerlund. p. 39).

Methods of rapid dissemination

Trusting sources of information

Source: www.youtube.com/watch?v=iV3cD9oYij4

Martin Lewis ostensibly recommends the viewers to invest in a pseudo-Elon Musk initiative (FB, July 2023).

Westerlund, Mika. "The Emergence of Deepfake Technology: A Review." *Technology Innovation Management Review* 9, no. 11 (November 2019): 39-52.

Question 6

-- / 0 pts

This is an optional question where you are welcome to share additional thoughts, questions, and requests.

Your Answer:

I am curious about how deepfakes would be different from impersonations.

Before the advent of digital or AI technologies, people would still run cons by having similar looking people pose as celebrities or have contact for 'friends' or 'family'. Since we already have legal processes for impersonations, why are deepfakes still not regulated?

On the other hand, if a business uses AI to animate historical characters and/or paintings, how would that play out? Would a talking Mona Lisa not be as problematic as a talking Einstein? And then how would we believe in the authenticity of 'recently discovered' or 'previously unknown' historical images and videos (and believe that they are not just clickbait)?

Finally, I'm also thinking of what deepfakes mean going forward. There was an interesting anecdote that I had heard - a medieval forger was so good at forging church documents that his forgeries ended up more valuable than the actual church documents. The history that we have from that time is a mix of real documents and fakes so well done that they are indistinguishable from the real. In the context of deepfakes, does this mean that going forward we propagate similar 'mixed' histories? And that we may leave for future generations a potentially false view of the world today?

Additional Comments:

Let's ponder

- "eventual value"
- "mixed history"
- "false view of the world today"

Post-class Active Learning survey: this class, in particular, evoked thoughtful responses.

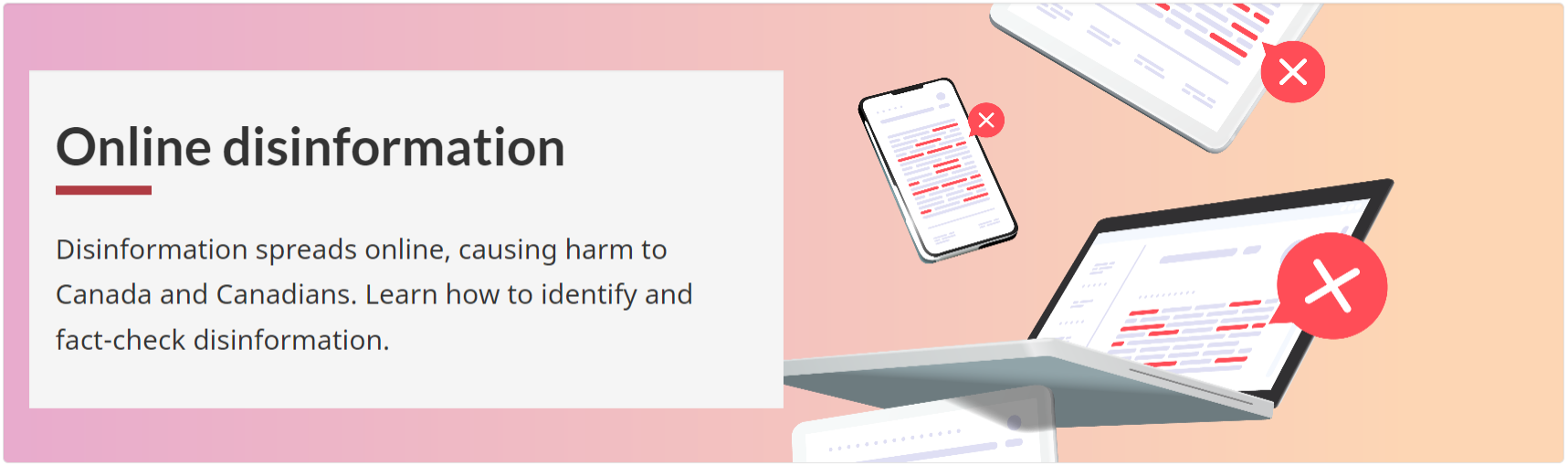
Français

Government of Canada | Gouvernement du Canada

Search Canada.ca

MENU

Canada.ca



Source: www.canada.ca/en/campaign/online-disinformation.html

Retrieved in November 2023

How disinformation undermines democracy?

- Disinformation about polarizing issues spreads and can lead to division in our society.
- Disinformation about persons and institutions can lead to mistrust in them.
- Disinformation influences the political landscape and can affect people's voting decisions.
- Over the long term, disinformation may even pose a threat to democracy itself.

Cyber threats to Canada’s democratic process

Figure 02: Short-, medium-, and long-term goals of state-sponsored cyber actors

Source: www.cyber.gc.ca

Updated in July 2021; retrieved in November 2023

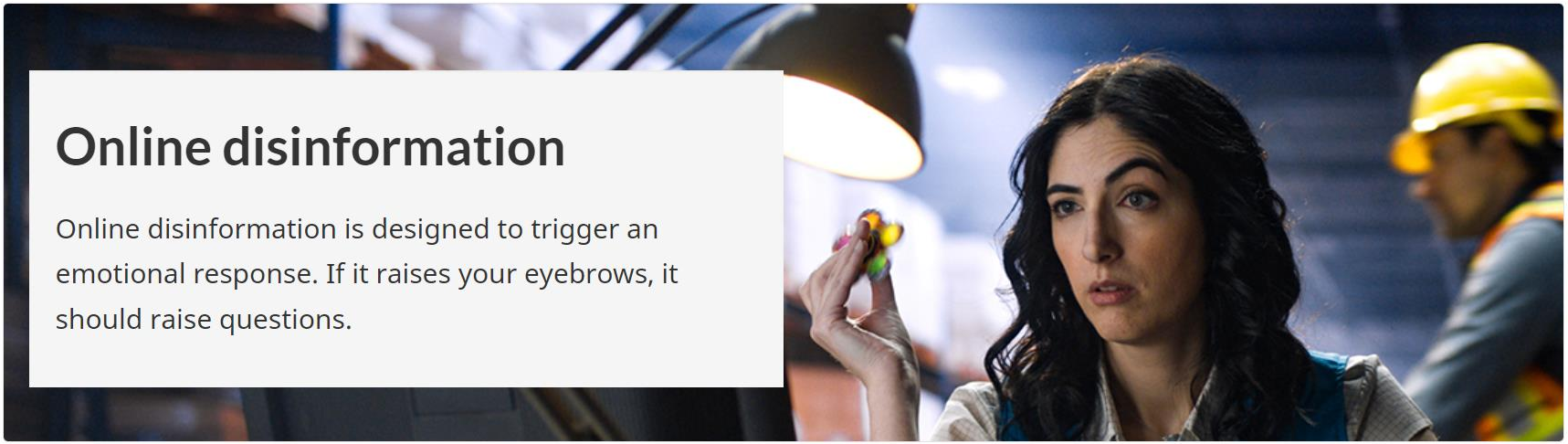
Français

Government of Canada | Gouvernement du Canada

Search Canada.ca

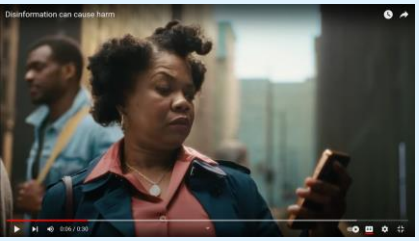
MENU

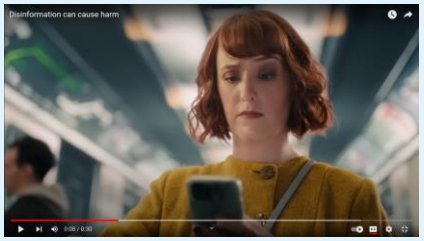
Canada.ca

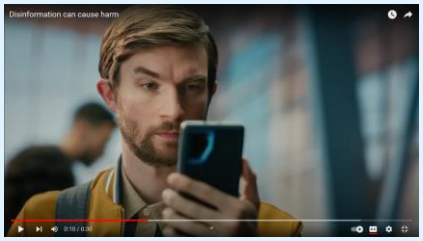


Source: https://www.canada.ca/en/campaign/online-disinformation.html

Updated in January; retried in June 2024







"If it raises your eyebrow, it should raise questions. Check the facts before you share online."

Authenticity becomes something to "believe" - implications on historical research

1. "Deepfakes pose a greater threat than "traditional" fake news because they are harder to spot and people are inclined to believe the fake is real..." (p.42).

Contemporary: I don't know what is real and what is not

2. "Further, people may even dismiss genuine footage as fake…, simply because they have become entrenched in the notion that anything they do not want to believe must be fake…" (p. 43).

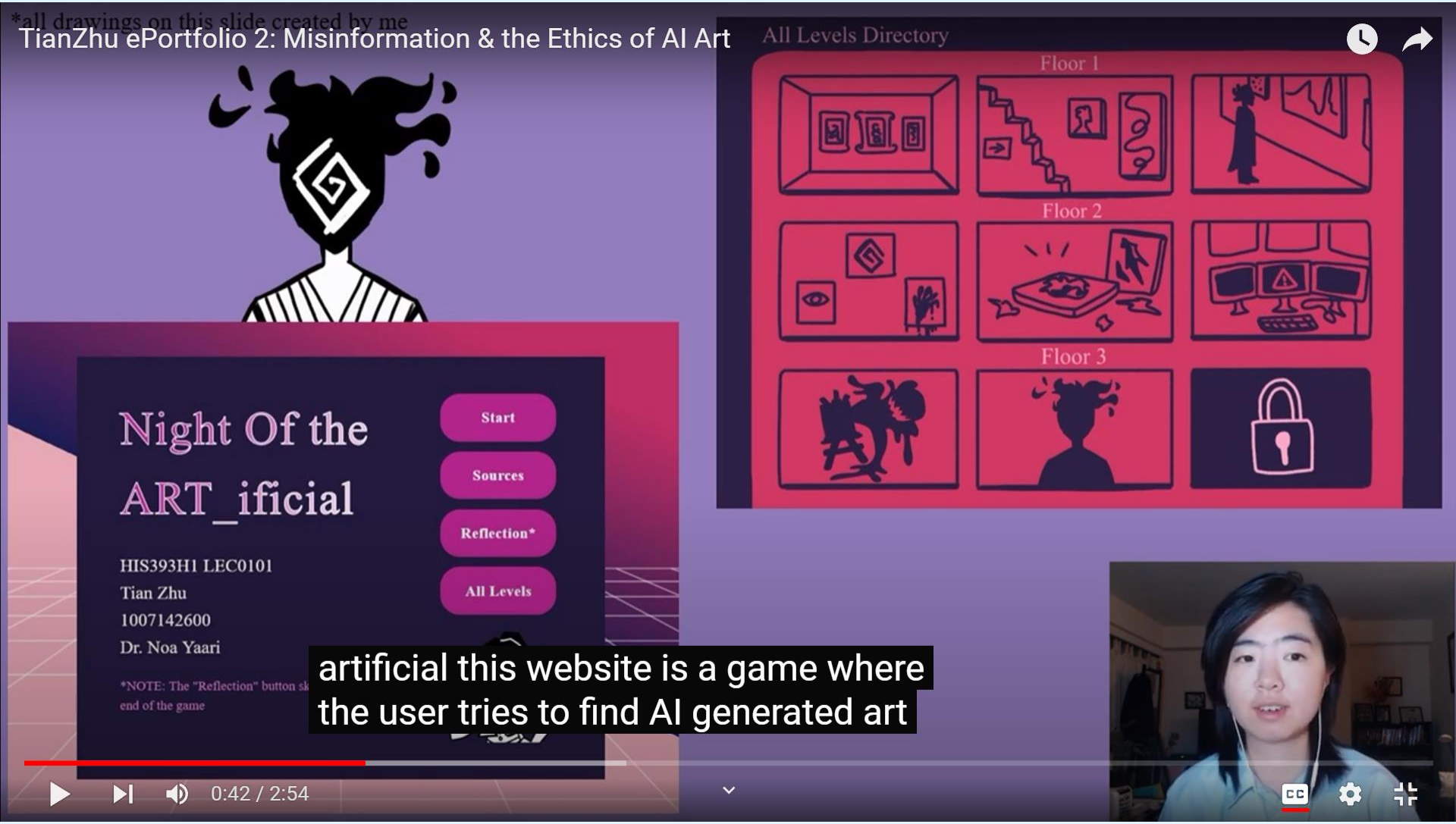
Contemporary: I don't believe it because it doesn't align with my worldview (and vice versa)

3. "Nonetheless, there will always be people who choose not to believe a verification tool, and rather still have a desire to consume and endorse fake media" (p. 46).

Contemporary: I mistrust verification tools and continue playing the game

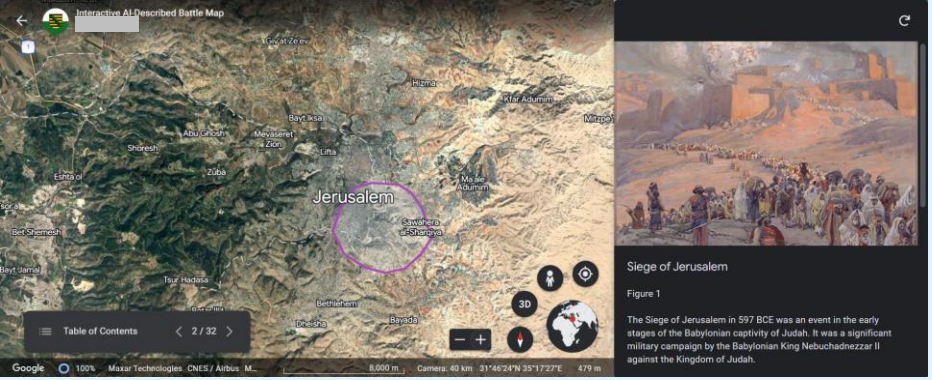
Westerlund, Mika. "The Emergence of Deepfake Technology: A Review." *Technology Innovation Management Review* 9, no. 11 (November 2019): 39-52.

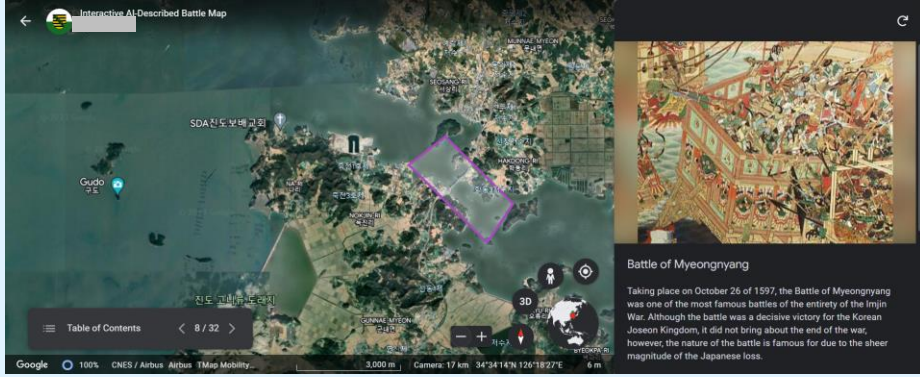
Business Model and Creative Project assignments to promote problem-solving in Digital History

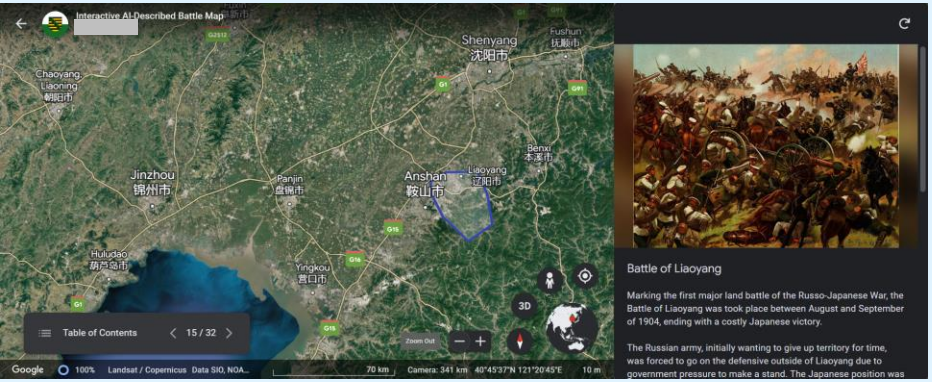


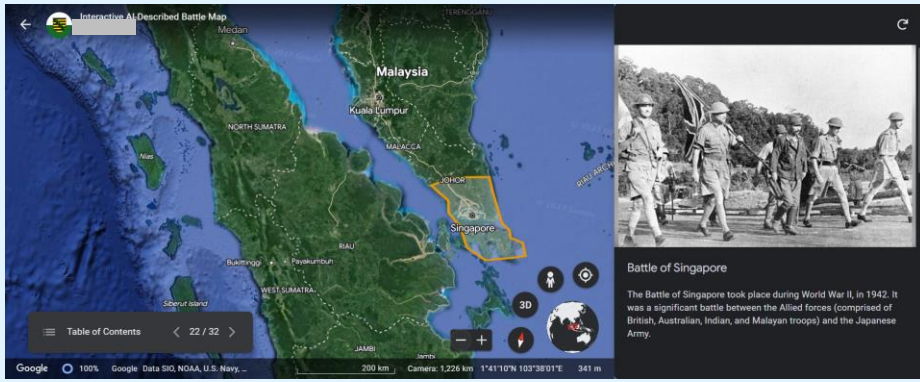
Source: https://www.youtube.com/watch?v=euJDjuuZ3dI

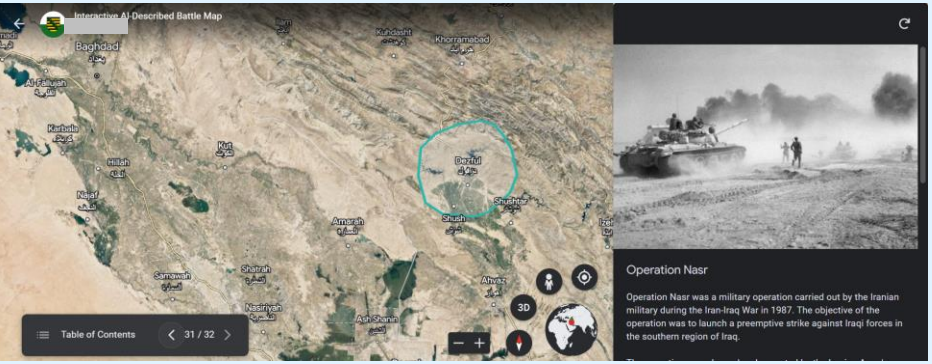
Tian Zhu, *Misinformation & the Ethics of AI Art*

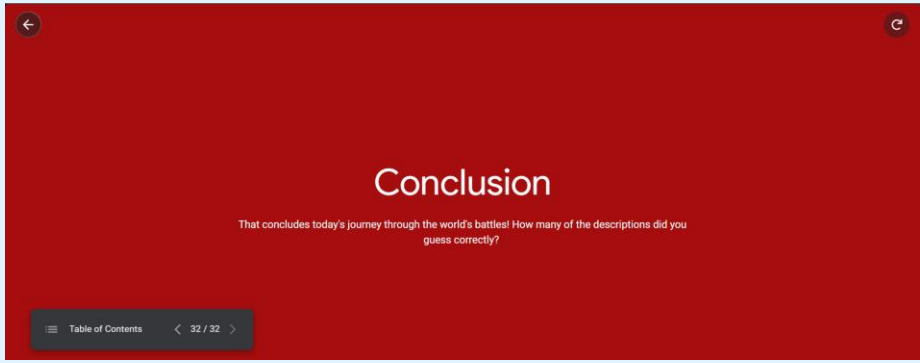












This project invites the user to determine whether the descriptions of the battles are human or AI-generated. Thorough this process “the user can better understand the challenges that AI poses to historical study, and how to identify these problems in the future.”

A. B. and D. G. *ChatGPT and History: A Battle Mapping Project*

Deepfake and historical research - closing with questions

Let's return to

- "**eventual value**" - of deepfakes
- "**mixed history**" - that includes both real and faked items
- "**false view of the world today**" - that will perpetuate disinformation

Eventual value of deepfakes

- Is deepfake a "regular" historical phenomenon or does it have a status of its own?
- What do deepfakes take from or contribute to historical research?

Mixed history

- Considering all eras are mixed, and the advancement of Gen AI, is our era more mixed than others?
- Will the future be even more mixed?

False view of the world today

- Will deepfake make historical research about our era radically different?
- Will historians be trained to navigate a mixed heritage and society with a deteriorating ability to tell the difference between true and false?