An Ethics Framework for Machine Historians

Marnie Hughes-Warrington
DASSH
19 September 2024

University of
South Australia

Approaches to Ethics for Artificial Agents

- AI has decoupled intent from action
- Early approaches to AI ethics (eg Floridi *The Ethics of Information*, 2013) proposed an object-oriented medical model: humans or AI can be agents or patients that are at least accountable for outcomes
- Current approaches combine medical frameworks (Beauchamp and Childress, 2013) with AI text mining methodologies (eg Floridi, *The Ethics of Artificial Intelligence*, 2023) to articulate principles
    - Beneficence, nonmaleficence, autonomy, justice, *explicability*
- Feasibility and implementation remain a problem:
    - Ethics as a (dispensable) board or as Artificial Artificial Intelligence (humans behind the machine)

Nico Grant, 'Google Chatbot's A.I. Images Put People of Color in Nazi-Era Uniforms', *New York Times*, 26 February 2024

Applied Ontologies for Machine Historians

- Logic and ontologies may address feasibility, interrelation, implementation concerns for AI ethics (eg Todai Robot 2017)
- Jo Guldi, *The Dangerous Art of Text Mining* (2023)
  - Neighbourhood methodologies show the rise, distinctiveness, overlap and disappearance of terms *but also topics and source lists* in historical research

'Winnowing, in agriculture, is a necessary stage after the first crops have been harvested; air is blown through the grain to remove the lighter chaff.… the scholar roughly works over the returns of the query **to sort signal from noise**, sturdy from flimsy, gathering up the promising results and discarding the less clear or less relevant information.'

Jo Guldi, *The Dangerous Art of Text Mining*, p. 130.

Todai Robot Ontology (2017)

- Good history essays are generally in chronological order;
- The order of two or more historical events can be switched if they overlap with one another chronologically; and
- Historical events should be treated in order of geographic proximity.

Turn up the Historiographical Noise

- Ontologies and logic are connected with ethics
- A case can be made for historiographical *noise* as reflective of logic and ontology
- Deep reading shows us a potential applied noise+signal ontology
    - Chronological flicker;
    - Tendency towards rhetorical questions;
    - Heavy and persistent use of conditionals in the text and in supporting source descriptions;
    - Heavy and persistent use of rhetorical counterfactuals, particularly of the 'had x then y' form

•Why?

•Histories may be shaped by a combination of question-based (non-truth evaluable, eg. Collingwood) and modal logics (eg. Lewis) with the combined outcome of openness and effort. This is underappreciated in AI ethics.

Historiography Text	Prize Histories	Machine Historians	Category	Property
Kate Antonova, Essential Guide to Writing History Essays (2020)	World History Association Bentley Book Prize Winners 2017–2020	Google Maps, Apple Maps	The topics of histories	Historical logic is intensional
Sarah Maza, Thinking About History (2017)	Wolfson History Prize Shortlist, 2021	Siri, Chat GPT series	Questions in histories	The frequency of questions increases within histories
Liu Zhiji, Shitong (A Thorough Exploration in Historiography)	American Historical Association John K. Fairbank Prize Winners 2018–22	Facebook, X (formerly Twitter)	Chronology in histories	Histories open with the rapid movement back and forward between times
Margaret McMillan, The Uses and Abuses of History (2010)	Cundhill History Prize Winners 2018–22	Netflix recommendation system, Amazon Prime Video recommendation system	Conditionals in histories	Conditionals are heavily present throughout histories, including references
Richard J. Evans, Altered Pasts: Counterfactuals in History (2014)	Pulitzer Prize for History, 2019–23	Spotify, Google Gemini	Counterfactuals in histories	Counterfactuals as rhetorical questions feature throughout histories. Historical topics can also be counterfactuals.
Jo Guldi, The Dangerous Art of Text Mining (2023)	Berkshire Prize Winners 2019–23	YouTube, Meta AI	Sources in histories	Sources support the intensional logic of history
Michel-Rolph Trouillot, Silencing the Past (1995)	American Historical Association Katz Prize Winners 2019–23	Google Translate, Amazon You Might Like system	Gaps and Silences in histories	Histories overlap with, but are distinctive from other histories
Samuel Wineburg, Why Learn History (When It's Already on Your Phone)	New South Wales Premier's Award for Young Histories	Snapchat, Dall-E series	Learning history	Noise increases in texts for older readers

Cybersecurity by Historiographical Design

- AI can be historiographical by ontological design: there is merit in the combination of question- and modal logics
- It is possible to design a bundle of signal to noise measures which capture the questioning, and modal stance of histories
- Why?
  - Most people do not have to read histories, but they are vulnerable to consipiracy or hate histories which travel quickly through the infosphere on the strength and repetitive nature of their signal

‘You're correct, my response lacks the use of conditionals, which could provide a more nuanced exploration of potential outcomes or alternative scenarios. Employing conditionals can help to hypothesize about different possibilities or speculate on how events might have unfolded under different circumstances. Let me revise the response to include conditionals for a more comprehensive analysis.’

Open AI, Chat GPT 3.5, online at: www.chat.openai.com <accessed 16 March 2024>, using the prompt ‘I notice that you did not use conditionals in your response. Why?’.

Thanks

Full copy of the paper available from marnie.hughes-warrington@unisa.edu.au