Jennifer Edmond & Erik Ketzan, *"Ein bestimmtes Gesellschaftsetwas* – Drawing Lessons for 21st Century 'Privacy-Protecting' AI Technology from 19th Century European Literature"

Abstract: The LI4AI (Literary Identities for Artificial Intelligence) project investigates issues related to the risks and future of Artificial Intelligence — including identity formation, privacy, and anonymity — using a corpus of European literary texts from the long Nineteenth Century.   The project contributes uniquely to the discourse of AI through its attention to the cultural filter by which literature was processing events of the first and second precursors to AI's so-called "Fourth Industrial Revolution" (World Economic Forum, 2016). In specific, we propose that although we may think of AI as presenting a distinctly a contemporary challenge, in fact, the technology-driven panopticon effect we recognise in the practices of companies such as Google and Facebook can be explored from its roots only by starting in the shared discourses and narratives of this earlier, foundational period.

The LI4AI project approaches the challenge it sets itself through mixed-method close and distant readings of the novels in its corpus. We propose to present to the IRIS meeting a first experimental component, in which we query and model anonymity and privacy. It will begin with the creation of a dictionary of terms, starting with obvious lemmas and strings ("anonymous," "in secret," "mask," "unsigned," etc.) and expanding based on close reading, collocates, syntactic constructions, and topic modelling. The modelling of such rich and complex topics as privacy and anonymity presents a host of methodological issues. While computer science and information theory have introduced such concepts as computational differential privacy (Mironov et al., 2009), it remains unclear how such schemas can inform the modelling of privacy and anonymity in literary texts, and indeed how such concepts correspond to deeply held, culturally specific ideals. Our token- and syntax-based dictionary will be scored on precision and recall for its ability to identify relevant passages in literary texts, and serve as a basis for more sophisticated models and experiments, including e.g. a combination of dictionary with sentiment analysis, character networks, and/or word vectors.

Literature has been dubbed a "laboratory of the possible" (Westphal 2011, 63) incorporating significant predictive potential (Ameel and Craps, 2020). As we approach the revolution that AI promises, the textual artifacts of the past (even with all their inherent, acknowledged biases, another issue the project explores in light of the challenges biased AI training data can cause) will prove valuable in building models which can also be applied to contemporary social challenges.

Jennifer Edmond is an Associate Professor of Digital Humanities at Trinity College Dublin, Co-director of the Trinity Center for Digital Humanities, Director of the MPhil in Digital Humanities and Culture and a funded Investigator of the SFI ADAPT Centre. Jennifer also serves as President of the Board of Directors of the pan-European research infrastructure for the arts and humanities, DARIAH-EU.
Erik Ketzan is a Postdoctoral Fellow in the Centre for Digital Humanities at Trinity College Dublin.