Annexure i: Artificial Intelligence and Intellectual Freedom

At the heart of many discussions around Artificial Intelligence (AI) are questions about its impact on intellectual freedom. These can span multiple dimensions - freedom to form and hold opinions without interference, freedom of expression, access to information, as well as broader individual self-determination, agency and autonomy. A 2018 report by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression provides a good overview of the potential impacts of AI technologies on the information environment.[1]

Naturally, there are ways in which AI can promote access to information and intellectual freedom. The Special Rapporteur's report cites, for example, the ability of AI to facilitate "broader and quicker sharing of information and ideas", and AI-powered translation methods can help individuals access information in more languages.

Similarly, an IFLA Trend Report chapter on Artificial Intelligence states that the Semantic Web – closely associated with AI – can positively impact access to information by improving search efficiency (while noting the potential adverse implications for privacy and censorship) and discusses the potential of AI-enabled multilingual voice translation to removing language barriers and improving accessibility online.

Conversely, there are also emerging concerns about the potential negative effects of AI on the information environment. In light of libraries' commitment to promoting intellectual freedom,[2] taking stock of these concerns can allow them to reflect on the role our institutions can play in mitigating potential downsides of AI.

Personalisation

One of the key AI applications that can impact the information environment is personalisation of individual online experiences. This point is often raised in relation to search engines and social media platforms, since both act as gatekeepers to large amounts of online information today.[3] Based on personal and non-personal data, AI embedded in such platforms chooses, ranks and organises the information an individual is exposed to.

As such, personalised content curation could have an effect on an individual's ability to seek and receive information and form opinions freely. A 2019 study by UNESCO highlights two concepts which are at times evoked to describe the potential adverse effects of AI-powered content personalisation. *Filter bubbles* limit the scope of information a user is exposed to by delivering content tailored to their interests, based on user characteristics and past engagement. Meanwhile, *echo chamber* refers to a

[1] Report of the United Nations Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression, 2018, https://undocs.org/A/73/348.

[2] As detailed, for example, in the IFLA Statement on Libraries and Intellectual Freedom, https://www.ifla.org/publications/ifla-statement-on-libraries-and-intellectual-freedom.

[3] For example, the 2017 "Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications" by the Council of Europe Committee of Experts on Internet Intermediaries points out a crucial role search engines play in shaping how an individual seeks, imparts and received information (https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5). Similarly, reports by Access Now ("Human Rights in the Age of Artificial Intelligence," https://www.accessnow.org/cms/assets/ uploads/2018/11/AI-and-Human-Rights.pdf) discusses AI in the contexts of both search engines and social media platforms.

phenomenon where exposure to a similar or repeated information can reinforce and strengthens a users' views or beliefs.[4]

Algorithmic curation can also lead to some users missing – or being shielded from – important social and political information.[5] For instance, the Special Rapporteur's 2018 report points out that AI-based targeting can have discriminatory effects and exclude people from information or opportunities – such as job or housing advertisements that exclude people on the basis of age, gender or ethnic background.

A 2019 Declaration by the Council of Europe Committee of Ministers specifically raises concerns about the 'manipulative capabilities of algorithmic processes'. The declaration highlights, for example, concerns about micro-targeting and algorithms' ability to influence a person's emotions, thought and decisions.[6] Concerns about potential impacts at a broader societal level have also been raised: for example, a 2019 study by the Council of Europe points out that such practices could interfere with the goal of enabling an inclusive and pluralistic public debate.[7]

As a result, the algorithms behind such curation could potentially impact freedom of opinion and expression with greater efficiency and on a larger scale than traditional media, yet are often opaque and non-transparent.[8]

Content moderation

Another AI application that could impact intellectual freedom is automated content moderation. Online platforms today are facing increasing pressure to address concerns about illegal or objectionable/harmful content. AI-powered moderation can take many forms: from deleting or blocking uploads of content identified as illegal or harmful, to de-prioritising such content, to automatic implementation of the right to be forgotten.[9]

The appeal of AI and algorithmic moderation here lies in its ability to work at a much larger scale and speed (as well as protecting human moderators from the psychological toll of the work).[10] Yet despite the advantages it offers, doubts are frequently expressed in relation to AI moderation. Many actors underline concerns over the current limitations of AI systems in recognising nuance, satire or irony, critical analysis, fair use, as well as broader cultural contexts in efforts to distinguish between legitimate speech and illegal or infringing activities.[11] These could lead to over-blocking and removal of

---

[4] Hu, X. et al., 2019, "Steering AI and Advanced ICTs for Knowledge Societies A Rights, Openness, Access, and Multi-stakeholder Perspective", UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0000372132. Notably, both concepts are currently examined in the academic field, and at present the evidence documenting the presence of such phenomena is mixed. Further academic inquiry into the suggested phenomena is needed.

[5] For example, an experiment by the Web Foundation in 2018 set up six identical Facebook profiles to examine the curatorial function of the algorithms. One of the findings showed that posts containing certain keywords – i.e. those pertaining to femicide and rape – did not surface for the six experimental user profiles, even when some of the posts had more "shares" and reactions than the stories that were displayed (http://webfoundation.org/docs/2018/04/WF_InvisibleCurationContent_Screen_AW.pdf).

[6] Council of Europe Committee of Ministers, 2019, "Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes", https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b.

[7] Yeung, K., 2019, "Responsibility and AI", the Council of Europe Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), https://rm.coe.int/responsability-and-ai-en/168097d9c5

[8] Report of the United Nations Special Rapporteur, "Promotion and protection of the right to freedom of opinion and expression".

[9] Brkan, M., "Freedom of Expression and Artificial Intelligence: On Personalisation, Disinformation and (Lack Of) Horizontal Effect of the Charter", 2019, https://ssrn.com/abstract=3354180.

[10] Hu, X., "Steering AI and Advanced ICTs".

[11] As mentioned, for example, in the "Study on the Human Rights Dimensions of Automated Data Processing Techniques", 2017, "Human Rights in the Age of Artificial Intelligence", 2018, or "Steering AI and advanced ICT", 2019.

legal content, potentially infringing on the principles of legality, legitimacy and proportionality of freedom of expression limitations.[12]

A related issue is the lack of transparency and oversight, the opacity of removal criteria, and the implied shift towards private companies determining what speech is permitted – which could also lead to over-blocking, since private companies might prefer to err on the side of caution (and indeed, may be incentivised to do so by legislation).[13]

Facts and evidence

An interesting aspect of AI's influence on the information environment is its potential impact on users' understanding of truth, facts and evidence. As the World Commission on the Ethics of Scientific Knowledge and Technology's "Preliminary Study on the Ethics of Artificial Intelligence" puts it,

"…AI brings specifically new challenges that are related to its interaction with human cognitive capacities. AI-based systems have implications for human understanding and expertise. Algorithms of social media and news sites can help to spread disinformation and have implications for the perceived meaning of 'facts' and 'truth', as well as for political interaction and engagement"[14]

A particular AI application that has received a lot of attention over the last years and was argued to have the ability to change people's relationship to 'truth' and evidence is deepfakes. These "synthetic' media files created with the help of neural networks and machine learning create depictions of people or events which are not real – but appear to be – and are argued to be able to 'challenge our sense of reality'.[15]

Crucially, the phenomenon of altered media is far from new. Deepfakes were preceded by Photoshop alterations, recontextualization, video speed alterations and other methods of audiovisual manipulation.[16] However, it is the use of AI to create deepfakes – potentially faster and in a way that is accessible to more groups than ever before – that has attracted significant amounts of attention and sensationalist coverage.

The heightened public awareness and concern that media can be altered by AI could be seen as one of the crucial impacts of deepfakes, as it can prompt people to call into question all the evidence they encounter, whether real or not.[17]

AI and Intellectual Freedom: literacy is key

Importantly, academic inquiry into the impacts of AI on access to information and freedom of opinion and expression is ongoing, and further research is needed to fully

---

[12] Further on the implications of AI solutions to online disinformation on freedom of expression: Marsden, C., and Meyer, T., 2019, "Regulating disinformation with artificial intelligence", Panel for the Future of Science and Technology European Science-Media Hub/European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624279/ EPRS_STU(2019)624279_EN.pdf.

[13] "Responsibility and AI".

[14] World Commission on the Ethics of Scientific Knowledge and Technology, 2019, "Preliminary study on the Ethics of Artificial Intelligence", https://unesdoc.unesco.org/ark:/48223/pf0000367823.

[15] Harper, C., 2018, "Machine Learning and the Library or: How I Learned to Stop Worrying and Love My Robot Overlords", Code4Lib Journal, https://www.academia.edu/37287826/Machine_Learning_and_the_Library_or_How_I_Learned_to_Stop_Worrying_and_Love _My_Robot_Overlords.

[16] Paris, B., Donovan, J., 2019, "Deepfakes and Cheap fakes: the manipulation of audio and visual evidence", Data & Society, https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1.pdf.

[17] Hao, K., 2019, "The biggest threat of deepfakes isn't the deepfakes themselves", MIT Technology Review, https://www.technologyreview.com/ s/614526/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves.

understand these phenomena. In the meantime, the example of deepfakes highlights the importance of efforts raising the public's media and information literacy.[18]

To consider critically whether a given piece of media is genuine, to know how algorithms influence the information one sees in search engines and social media platforms, to be more confident in an increasingly AI-mitigated information flow – digital, media and information literacy is an important part of the solution.[19] Literacy efforts offer a way to tackle such concerns without compromising freedom of expression,[20] or intellectual freedom at large.

As further discussed in Annexure ii, libraries are inclusive spaces that specialise in information literacy – as such, they can be well-positioned to deliver such training to help people navigate the new information landscape. This could, in fact, be of the key roles libraries can play in ensuring that AI benefits society at large.[21]

Artificial Intelligence in the Library: Privacy Considerations

Privacy considerations

Another key ethical consideration to keep in mind with regard to implementation and use of AI is privacy. The training of AI systems can require large amounts of data, as can AI decision-making, the calibration of systems or the refinement or creation of feedback mechanisms.[22] This can range from non-personal data to sensitive or personally identifiable information; with most use-cases falling somewhere in the middle of the continuum (e.g. relying on information inferred from personal data).[23] Such uses already imply data protection issues, given the principle that people should have a say over how their information is used.

Moreover, outputs of AI-driven analysis can also reveal personal or sensitive information through crossing data sources, or through making it possible to de-anonymise personal data. This raises questions about an individual's right to privacy and data protection.

This can also concern AI applications in a library setting. For instance, possible user-facing AI use-cases are numerous: "Due to algorithmization many possibilities arise to connect (potential) users to collections, to personalise user interactions and to optimise various aspects ranging from space usage (for physical parts of the collection) to advanced retrieval methods based on complex user demands."[24]

Some of these interactions can entail important privacy implications for a library user. A 2019 study on use-cases and implications of AI in academic libraries, for example, saw a number of interviewed participants mention the possible privacy implications.[25]

[18] Wagner, T., Blewer, A., 2019, "The Word Real Is No Longer Real": Deepfakes, Gender, and the Challenges of AI-Altered Video", Open Information Science 2019; 3: 32–46, https://www.degruyter.com/downloadpdf/j/opis.2019.3.issue-1/ opis-2019-0003/opis-2019-0003.pdf

[19] For example, literacy is highlighted as an important way to tackle the potential challenges of AI for freedom of opinion and individual autonomy in the "Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes".

[20] Marsden, Meyer, "Regulating disinformation with artificial intelligence".

[21] For instance, the idea of libraries teaching algorithmic/information literacy had been expressed by several participants at the 2018 CFLA-FCAB National Forum discussion on Artificial Intelligence: http://cfla-fcab.ca/wp-content/uploads/2018/05/AI_table_notes.pdf

[22] "Human Rights in the Age of Artificial Intelligence". See also Campolo, A. et al., 2017, "AI Now 2017 Report", AI Now, https://ainowinstitute.org/AI_Now_2017_Report.pdf.

[23] Report of the United Nations Special Rapporteur, "Promotion and protection of the right to freedom of opinion and expression".

[24] Van Otterlo, M., 2018, "Gatekeeping Algorithms with Human Ethical Bias: the ethics of algorithms in archives, libraries and society", https://arxiv.org/abs/1801.01705.

[25] Cox, A.M., Pinfield, S. and Rutter, S, 2019, "The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries", Library Hi Tech, 37 (3). pp. 418-435.

These can include, for instance, the collection of data about activity in the physical library space, or a user's borrowing and reading behaviour in order to feed AI systems.

In such cases, it is important to ensure both basic data privacy/security and the intellectual privacy of library users.[26] In fact, privacy considerations are fundamentally related to intellectual freedom matters discussed in the earlier section. Privacy violations can have a chilling effect on freedom of expression,[27] and the knowledge that collected data could be re-used by AI systems in other contexts (without awareness and/or control by the data subject) can prompt people to refrain from participating in some activities that entail data collection.[28]

What libraries can do

As the 2015 IFLA Statement on Privacy in the Library Environment discusses, there are measures libraries can and should take to protect patrons' data.[29] In the context of AI, this can include ensuring that any experiments with AI applications which modify or nudge library user behaviour are carried out on an informed opt-in basis,[30] or carefully reviewing library choices of third-party AI applications to see how they impact users' privacy.[31] More broadly, AI applications should not unduly compromise user privacy.

The concept of consent for personal data processing is also important here. As a report by a Council of Europe Consultative Committee highlights, the obscurity and complexity of AI algorithms can make it challenging to have a user's informed consent for data processing. Suggested remedies include increased transparency and carrying out risk assessments, as well as considering the merits of introducing flexible forms of consent (e.g. dynamic consent).[32]

Artificial Intelligence in the Library: Bias, Transparency and Explainability

Addressing AI bias concerns

A key ethical consideration most frequently raised in the context of AI development and use is the possibility of AI bias leading to unfair or discriminatory outcomes.[33] In the information environment, for instance, an often-cited example is the autocomplete search engine function that returned suggestions laden with harmful stereotypes.[34]

Such bias can be introduced in many ways and at different phases of AI development and use. This can occur when framing and formalising the problem an AI model is intended to address, or when labelling examples or defining and selecting the attributes

---

[26] Van Otterlo, M., and Warnaar, M., 2017, "Towards Optimizing the Public Library: Indoor Localization in Semi-Open Spaces and Beyond", http://martijnvanotterlo.nl/benelearn2017-vOtterlo-Warnaar-CR.pdf.
[27] "Human Rights in the Age of Artificial Intelligence".
[28] "A study of the implications of advanced digital technologies".
[29] "IFLA Statement on Privacy in the Library Environment", 2015, https://www.ifla.org/files/assets/hq/news/documents/ifla-statement-on-privacy-in-the-library-environment.pdf.
[30] Van Otterlo, M., 2016, "Project BLIIPS: Making the Physical Public Library More Intelligent through Artificial Intelligence", Qualitative and Quantitative Methods in Libraries (QQML) 5: 287-300, http://www.qqml.net/papers/June_2016_Issue/522QQML_Journal_2016_Van_Otterlo_287-300.pdf.
[31] Harper, C., 2018, "Machine Learning and the Library".
[32] Mantelero, A., 2019, "Artificial Intelligence and Data Protection: Challenges and Possible Remedies", Council of Europe Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6.
[33] E.g. as explained in Laterono, M., 2018, "Governing Artificial Intelligence: Upholding Human Rights & Dignity", Data and Society, https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf;
[34] As described in Dr. Safyia Umoja Noble's book "Algorithms of Oppression", 2018.

an AI system should consider.[35] Biases can arise as a result of the training data an AI was exposed to: datasets reflecting historic or existing societal inequalities, or datasets that under- or over- represent disadvantaged groups may lead to decisions that replicate these inequalities. Even when a dataset does not include protected characteristics like gender or race, these could be inferred from other data in the set, and lead to discriminatory results.[36]

These concerns can be relevant for AI applications in the library sector as well. Contributors to a 2019 OCLC Primer expressed concerns that the growing use of algorithms in the library setting could reinforce biases that negatively impact library staff, patrons, or society at large. The Primer offers a range of recommendations to manage potential bias and commit to responsible options for AI use in cultural heritage institutions. These include, for example, initiating "practice exchanges" and holding symposia on approaches to managing bias when adopting machine learning (ML) and AI technologies; developing auditing approaches; and forming committees to guide responsible engagement with AI and ML.[37]

For externally-sourced tools, this means that it is crucial for libraries to know how an outside vendor has carried out the training of the AI.[38] For AI projects originating in the library, this means carefully reflecting on the potential ethical implications: for example, for a project aiming to use ML to generate subject headings, it would be important to consider existing biases in subject headings developed by cataloguers.[39]

Finally, a greater focus on explainability and transparency in library ML-powered systems (despite significant technological challenges and possible performance trade-offs) can offer important benefits to detecting and addressing bias.[40]

Librarians helping develop ethical AI

Finally, it has been suggested that librarians – in particular research and data librarians – can help shape AI development "for good". The Association of Research Libraries 2019 issue on the ethics of AI, for example, mentions that research librarians' expertise in knowledge management and research support can help them participate in public policy debates on AI governance towards the public good.[41]

They could help researchers manage and work with their data in replicable and consistent ways, taking account of data quality and curation issues; as well as offer guidance on privacy and ethical use of personal information policies; and help students develop such research skills.[42]

---

[35] Hao, K., 2019, "This is how AI bias really happens—and why it's so hard to fix", MIT Technology Review, https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/;

[36] Borgesius, F. Z., 2018, "Discrimination, artificial intelligence, and algorithmic decision-making", Directorate General of Democracy - Council of Europe, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.

[37] Padilla, T., 2019, "Responsible Operations: Data Science, Machine Learning, and AI in Libraries", OCLC Position Paper, https://www.oclc.org/content/dam/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.pdf.

[38] Griffey, J., 2019, "Chapter 1: Introduction," Library Technology Reports 55, no. 1: 5–9, https://journals.ala.org/index.php/ltr/article/view /6908/9300. Notably, the author expressed doubt that such scrutiny will occur in practice, therefore encouraging libraries to be involved in development of their own tools.

[39] Boman, C., 2019., "Chapter 4. An Exploration of Machine Learning in Libraries", Library Technology Reports 55, no. 1: 21-25, https://journals.ala.org/index.php/ltr/article/view/6911/9303.

[40] Koene, A., et al., 2019, "governance framework for algorithmic accountability and transparency", European Parliament Panel for the Future of Science and Technology, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf.

[41] Ridley, M., 2019, "Explainable Artificial Intelligence." *Research Library Issues*, no. 299 (2019): 28–46, https://doi.org/10.29242/rli.299.3.

[42] Henry, G., 2019, "Research Librarians as Guides and Navigators for AI Policies at Universities." *Research Library Issues*, no. 299 (2019): 47–65, https://doi.org/10.29242/rli.299.4.