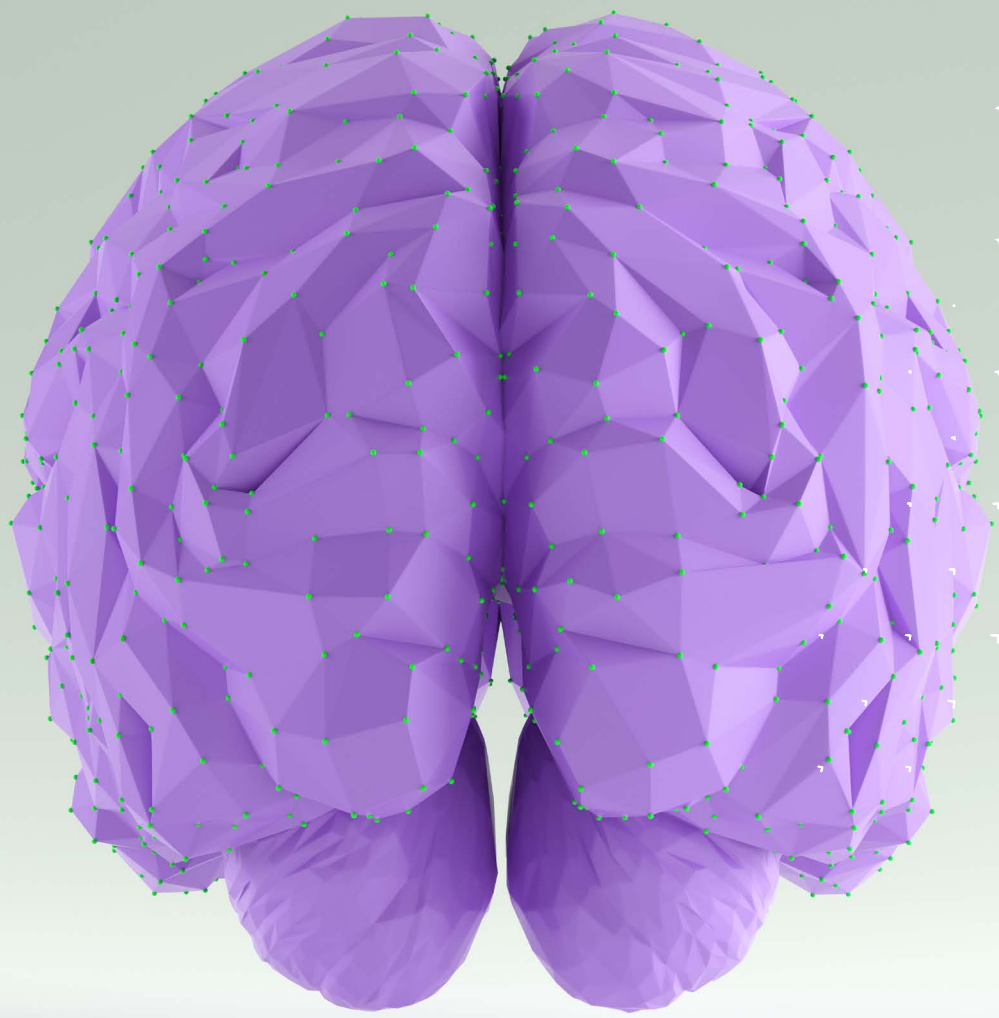
STM Advancing trusted research

Ref. Ares(2021)4884972 - 30/07/2021

AI Ethics in Scholarly Communication

STM Best Practice Principles for Ethical, Trustworthy and Human-centric AI



Contents

1 / Introduction

Artificial Intelligence (AI) is everywhere in today's society, the news and popular culture. The academic publishing sector is no exception. Although AI is still an emerging technology, many publishers are already productively employing AI and contributing to its development in many ways. For its promise to be fulfilled and truly improve research, science, technology, medicine and broader society, AI has to be grounded in the values of trust and integrity fundamental to scholarly communication. Several papers have addressed and discussed the legal and ethical issues of AI.[1] Building on these general principles, STM (the International Association for Scientific, Technical and Medical Publishers) felt that it would be worthwhile to delve into AI and in 2019 the association formed a working group to explore the specific perspectives that the STM community brings to the issues raised.[2] This White Paper brings together the working group's current thinking on how STM publishers contribute to the ethical and trustworthy development, deployment, and application of artificial intelligence. The paper is not intended to be exhaustive; rather it aspires to contribute to the ongoing discussion on how to move forward with this technology.

With advances in computing power, big data and algorithms have made AI an increasingly ubiquitous reality. The form and complexity of AI varies widely, and undoubtedly new forms will emerge over time. As a result, this paper will focus on high level principles rather than specific implementations.

AI has been defined as machine-based systems, operating on a large scale with varying levels of autonomy, that can "make predictions, recommendations, or decisions influencing real or virtual environments."[3] AI is relevant in any context where large volumes of data and information are processed. Today, it is a strategic technology that offers many potential benefits for citizens, the public good and the economy at large,[4] provided it is human-centric, ethical, and respects fundamental rights and values.

With advances in computing power, Big Data and algorithms have made AI an increasingly ubiquitous reality. Given that the form and complexity of AI applications varies widely, and undoubtedly new forms will emerge over time, this paper focuses on high level principles rather than specific implementations. However, the potential of AI for science is certainly already promising. Science has evolved through technological innovation from its early beginnings. From being primarily observational in ancient times, the 17th and 18th-century inventions of the microscope, telescope and other tools have made it increasingly experimental. The introduction of the computer in the 20th century ushered in a new era and has since become an indispensable tool for academics in nearly all aspects of their workflow, sparking a new wave of computational science. The computer allows scientists to collect, automatically share and analyze huge amounts of data, making science more data-focused in recent decades. Fueled by the availability of digital content including data, articles, and books, AI now has the potential to truly revolutionize science.[5]

[1] See e.g., https://www.oecd.org/going-digital/ai/principles/,

https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai,

https://dataresponsibly.github.io/

[2] Duncan Campbell (Wiley), Elizabeth Crossick (RELX/Elsevier), Victoria Gardner (Taylor & Francis), John Ochs (ACS), Henning Schoenenberger (Springer Nature), David Weinreich, Claudia Russo and Joris van Rossum (STM).

The examples of AI implementations at publishers throughout this white paper were written by Sonja Krane (ACS), Scott Dineen (OSA), Mathias Astell (Hindawi), Marcel Karnstedt-Hulpuş (Springer Nature), and Jabe Wilson (Elsevier).

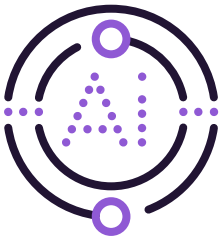
[3] Organisation for Economic Co-operation and Development (2019), *Recommendation of the Council on Artificial Intelligence*. Retrieved from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[4] See e.g., https://www.pwc.co.uk/economic-services/assets/macroeconomic-impact-of-ai-technical-report-feb-18.pdf

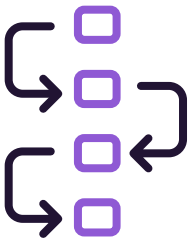
[5] See e.g., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Tony Hey, Stewart Tansley, Kristin Tolle, Published by Microsoft Research, October 2009.

Publishers are involved with AI in three areas:

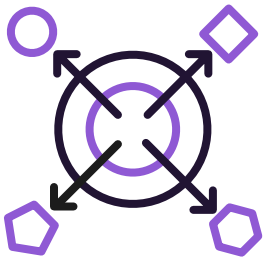
1 / Data providers



2 / Supporting internal workflows and services



3 / External-facing tools and services



If it is applied in a trustworthy, ethical, and human-centric way, AI holds the promise for so-called 'smart science', not just testing hypotheses against vast amounts of data but also creating new ones, developing new theories, exploring new connections and determining hitherto unknown causes.

STM publishers have long been the partners of academic scholars and scientists, since the scientific revolution of the 16th century, by connecting researchers, their research and the wider world. Publishers are continually innovating to add value to an increasingly digital and interconnected environment. AI continues to deepen and broaden that partnership. Currently, publishers are involved with AI in three broad areas.

First, publishers are key providers of information and data on which AI is run. Relevant, high-quality input and training data for AI developers and systems form one of the key ingredients for high-quality, trustworthy and ethical outputs. Providing this corpus of data in required digital formats is a core expertise of publishers. By validating, normalizing, tagging and enriching content, delivering material in robust, interoperable and globally consistent formats, and creating domain-specific ontologies, publishers ensure that information is a trustworthy high-quality input source with tremendous potential for use by AI systems across a broad range of applications.

Second, publishers increasingly use AI – either developed in-house or supplied by third parties – to support internal workflows and services for authors, editors, and reviewers. For example, AI is used in recommending journals to authors based on sections of manuscripts, streamlining submissions by carrying out technical and language checks, and identifying suitable peer reviewers. Many publishers use AI to detect plagiarism, spotting suspicious patterns in content. New attempts include using AI to identify image and data manipulation. Novel applications also include extracting automatically the most important facts, entities, and relations from scientific papers.

Third, publishers deploy AI in external-facing tools and services, using it to classify content, recommend related content to readers, and to author new content by bringing together related information from disparate sources. Fueled by AI, publishers are increasingly serving as providers of analytics and insight,[6] by providing insights into research trends, for instance, or as input for R&D and by identifying targets for drug development.[7]

STM publishers are both users and producers of data, in different roles for the latter. For example, STM publishers are engaged in publishing primary content (journal articles and books), creating databases, and facilitating links between journal articles and research data stored elsewhere. Whether and what rights are involved depends on the context and specific role STM members fulfil.

Because AI is an emerging technology, the nature of the technology and our ideas on how to engage with it will inevitably shift and change, which makes it inevitable that this work will need to be updated. Regardless of this likelihood, this White Paper is an attempt to lay a solid foundation for an ethical and trustworthy implementation of AI in STM publishing and in scholarly communications at large.

[6] Radical Reinvention - Outsell's Annual Information Industry Outlook 2021. https://www.outsellinc.com/product/the-outsell-information-industry-outlook-2021/

[7] In their traditional role, publishers are also helping to advance the field of AI research by publishing articles, books and launching subject-specific journals and databases.

2 / Legal and policy framework

STM proposes that the further development of AI be guided by and grounded in clear legal standards and sound ethical principles. As AI technologies continuously evolve, newly introduced legislative tools bear the risk of being overly inflexible, possibly jeopardizing the innovative processes they are meant to support. This may lead to unintended and perhaps harmful consequences. This is why, when considering the development of new laws, policymakers may wish to determine first if existing legislation is adequate to address AI regulatory needs.

Many regions and countries are considering the need to either regulate AI or to adapt current regulation. The first region to propose action in this area is the EU with the Proposal for a Regulation on a European approach for Artificial Intelligence, released on April 21 2021.[8] Where new AI policy may be necessary, it should be underpinned by evidence and developed through broad consultations, with defined outcomes including a solid evaluation of the expected impacts and clear policy options. STM backs the use of external advisory committees providing expertise and experience in support of legislators for AI policy formulation, for instance providing guidance on how data governance and management systems may determine how key inputs may be used, generated, stored, and potentially re-used.

An intellectual property (IP) policy framework is recommended that recognizes IP as a fundamental right, and continues to incentivize investment in high-quality content, datasets, metadata, and curated items as well as databases that can securely be ingested into or associated with AI applications. This would help ensure an environment that fosters the development of innovative AI tools and respects the rights held in content used to generate or improve those systems. This includes funding and incentives for licensing conditions that are both flexible and adaptable, leading in the long term to higher quality, assured provenance, and thus, more trustworthy AI technologies supported by sustainable business models.

Legal certainty is important in fostering trust and investment in AI development, design, and operation, and it should actively be embedded in policy frameworks. A natural consequence of this principle is that any AI project needs to comply with applicable laws, including laws regarding:

- Copyright and other intellectual property,
- personal data,
- freedom of information,
- equality and discrimination,

ACS Publications
Most Trusted. Most Cited. Most Read.

Most stages of the editorial peer-review process could be enhanced through the application of artificial intelligence. The American Chemical Society (ACS) publications seek to support their author, editor, reviewer, and reader communities by leveraging technology to assist with journal selection, reviewer recommendation, related content and other constructive activities. As an example, manuscript transfer helps authors publish with ACS and reduces the burden on editors and reviewers; ACS AI tools suggest transfer destination journals within the portfolio based on semantic analysis and publishing history. The tool was recently linked with ACS' peer review system for ease of use.

[8] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

- sharing and re-use of private and public sector data,
- sector specific data (e.g., health, health research),
- and the production of statistics for decision-making.

STM recommends that any AI-enabling policy framework fosters the development of community-based standards and best practices on an agreed time frame that meets the needs and goals of the key stakeholders. Where possible, policy should build on existing initiatives to ensure alignment, minimize burdens on stakeholders and avoid duplicating efforts. STM suggests promoting positive and clearly understandable examples of AI applications to the public to foster broad understanding and support, and to underpin sustainable long-term public and private investment in AI.

Ethical behavior in AI design and deployment can be achieved with the use of various regulatory tools, including self- and co-regulation by industry. Dissemination of best practices will also help with the ambition of achieving quality in AI. As mentioned, policymakers may want to include external advisory committees with a broad range of expertise and experience in AI to review policy, and potentially, individual AI processes, paving the way to achieving an appropriate instrument. Making use of expertise developed in the industry from the outset is an important factor in ensuring uptake of quality AI systems. Building on existing initiatives, upholding the quality of existing systems, and adapting those for the AI setting will also help avoid creating new redundant workstreams.

3 / STM best practice principles for ethical and trustworthy AI

STM's best practice principles for ethical and trustworthy AI may be grouped in five categories:

1. Transparency and Accountability

Transparency and accountability of AI can most clearly be achieved on the level of the data used in AI training and input, as well as in the use of AI technology in publisher's tools, processes and services. Publishers are both committed and uniquely positioned to support transparency in data – providing information on its provenance and ensuring data is available in a structured and consistent format – while clarity and transparency around the use of IP and copyright materials is required. AI use should be particularly transparent in cases where it is used as a decision tool, especially to those directly affected. Questions around establishing accountability structures are and should be subject to continuous mediation between all stakeholders in science and research. Publishers encourage working with other actors to adapt standards where necessary.

2. Quality and Integrity

Ensuring quality and integrity is fundamental to establishing trust in the application of AI tools and services. These values should be at the heart of the AI lifecycle, from the design and building of algorithms, to inputs used to train AI tools and services, to those used in the practical application of AI. Publishers play a vital role in supporting the quality of the AI ecosystem as suppliers of content that users can trust. An appropriate IP framework is essential to support sustainable services.

3. Privacy and Security

Privacy and security stand at the basis of ethical AI that many countries have formulated as a legal requirement. Several principles that focus on data protection, data privacy and security can and should be used to respect and uphold privacy rights, data protection and ensure the security of datasets used in training or operating AI systems.

4. Fairness

AI is based on identifying patterns in existing data, which holds the risk of historical bias. This may result in discrimination and prevent the emergence of novel ideas and theories. To avoid this, data selection and the application of AI must be carefully analyzed, planned, reviewed, and continuously monitored. Feedback mechanisms should be developed that can report cases or concerns of bias.

5. Sustainable Development

The multi-disciplinary nature of AI systems makes them ideally positioned to address areas of global concern, such as the United Nations Sustainable Development Goals. It also provides opportunities

Hindawi

The requirement to publish in the English language provides hurdles for non-English native speakers to overcome, not just as part of submission but all the way through the publishing process. As a publisher with a large proportion of non-English native speaking authors, Hindawi has sought to reduce this burden on authors by freely providing an AI language assessment tool as part of the submission process. Since implementing this tool, they have found a reduction in the number of references to language issues in peer review reports and a reduction in turnaround times during peer review. This use of AI has proven to be a valuable service to its users, not only creating a better experience for authors, editors and reviewers but also enabling a smoother and more effective pipeline for processing articles.

for greater efficacy in public and private organizations to achieve greater ecological sustainability and responsibility. AI systems bear the promise to benefit all humans, including future generations. Funding and other incentives for suppliers of high-quality input data, such as the publications and databases created by publishers, can help to extract important actionable knowledge.

3.1 / Transparency and Accountability

Accountability lies at the heart of research. Scientific progress is not linear; rather, it acts like a conversation, with discourse among scholars often diverting down blind alleys; with researchers working in the same areas as their peers; and by building on earlier endeavors ("standing on the shoulders of giants"). The key to accountability in science lies in transparency and publishers play an integral role in this process. Publishing organizations make significant efforts and investments to achieve transparency in research though publications, peer reviewed for quality and correctness, linking together outputs (e.g., data, software, journal articles and books), adding metadata, and applying persistent identifiers such as DOIs. The current ecosystem of scholarly communications allows users to interact not only with the final written narrative, but also the wealth of source material that researchers have used and cited, as well as supporting data and artifacts created in the research process.

With AI, the problem of determining accountability has become yet more complex. In some areas of AI, such as Deep Learning, algorithms generate new algorithms. Therefore, it is often not possible to understand, let alone explain (e.g., by reproduction or reverse engineering), how new algorithms and decision-rules were developed or the means by which a deep neural network came independently to a specific set of results. Unlike mainstream computing, where a given series of "if, then ... else" statements leads to an inevitable conclusion, machine-learning systems make probabilistic assessments based upon processing a myriad of iterations and reiterations of input data. This can make sophisticated AI systems and algorithmic processes increasingly complex and opaque by design. Thus there is often no clear pathway to trace back when looking to explain a specific decision.

For many commentators, the fundamental requirement of AI systems is that their use be transparent: that consumers and citizens are made aware when algorithms and machine-learning tools are in operation. The stronger the role AI plays in making decisions or recommendations, the greater the need for this transparency.

Transparency in training and input data is seen as another crucial element of responsible AI when outcomes are sensitive or controversial. In these cases, questions arise about the underlying datasets used, how the data was sourced, and whether it was aggregated and adapted to make it usable in this context. For example, it has been well-documented that some AI systems for facial analysis can display significant gender and racial bias, depending on the underlying training data used, and assumptions made during the development process.[9] By advocating and checking transparency in the scientific process, STM publishers intend to play a role in achieving better transparency in AI.

[9] https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/

3.1.1. STM best practice principles

- Transparency – on what underlying datasets have been used and how the data has been sourced, aggregated and adapted to make it usable – is crucial to be able to understand and mitigate issues with AI outputs. Publishers are custodians of greater data transparency by providing metadata on the origin, provenance and validation process of data. Where possible, this metadata is enhanced by consistent tagging enrichments in robust, globally consistent formats and domain-specific ontologies.
- When an AI-driven system is used as a decision tool, especially in cases of accusations or suspicions of unsound scientific practice, or even fraud (e.g., image manipulation), or AI is responsible for certain outputs without human intervention (e.g., textual summaries or books), it should be clearly indicated that these outcomes were based on AI.
- Where AI is deployed in the peer review process, it is important to communicate this transparently to all involved in the process (from authors to peer reviewers to readers of articles). In general, when AI is part of the research lifecycle, pre-existing standards and best practice conventions should apply to AI.
- Providing full accountability and transparency is desirable when publishers use AI in the publishing process or in customer-facing tools and services in an advisory or recommending capacity (e.g., to present editors with a list of possible reviewers; authors with a list of journal titles to submit their manuscripts to; readers with a list of recommended articles to read), or in back-office processes (e.g., technical/language checks on incoming manuscripts).
- Clarity and transparency are required in the use of IP and copyright, and as part of any liability regime. AI systems can use huge volumes of copyright materials in the training process and as part of any commercial deployment, therefore transparency obligations may be necessary to enable rights holders to trace copyright infringements in content ingested by AI systems.
- Accountability in research and science is a responsibility shared between key stakeholders, including researchers, funders, policy makers, and publishers. Publishers are committed to work with other stakeholders to establish accountability standards and adapt standards where necessary.

OSA Publishing

Like many journal publishers, the Optical Society (OSA) is always seeking better ways to identify appropriate reviewers for manuscript submissions across a broad range of topics. To that end it developed an AI tool that identifies reviewer candidates based on recent publication history across OSA's journal portfolio. The tool helps identify potential reviewers based not just on who editors know but on the technical expertise derived from AI analysis of authors' recent journal publications. As a result, OSA journal editors are able to see reviewer candidates based on the similarity process along with any peer-review history we have for the authors of matching papers including responsiveness, current obligations, and any potential conflicts.

3.2 / Quality and Integrity

When evaluating AI applications for quality and integrity, all aspects of the system need to be considered – from design, implementation, inputs, to processing algorithms and eventually outputs. While quality is ultimately judged by the value and reliability of the outputs, it should be possible to trace any concerns back to any one of these steps. The sometimes opaque nature of AI technology means that even with the most careful evaluation, the operation of some advanced AI systems may be unclear and can only be evaluated in terms of the output they produce. This presents challenges, but these can be addressed through rigorous testing and evaluation, transparency, and feedback loops. It is important to address the quality and integrity of the entire AI cycle while adhering to commonly shared ethical standards and best practice in the design and deployment of AI.

Quality outputs depend on the accuracy and reliability of algorithms and the code that implements them and also, critically, on the quality of the data input. A process can only be as good or unbiased as the input used to teach the system. Better data improves the efficacy of an AI tool or service. AI cannot be trained on mere facts alone; the context in which those facts are used is also important. The availability and accessibility of high-quality training data is vital for empowering AI developers with the materials needed to achieve AI quality and integrity. STM publishers are at the forefront of digital innovation, providing well-formatted digital data and information, tagging and enriching content and creating ontologies. This kind of structured and enriched information, supported by interoperable standards and persistent identifiers, is essential for the integrity of AI systems. It means that publishers can help ensure transparency, foster integrity, and drive quality in the research and publishing ecosystem, including the AI tools and services deployed in this ecosystem. Curating the underlying information will help to make the data more useful and the resulting AI tools more trustworthy.

The owners of content, including datasets protected as copyright works or as protected subject-matter under a related right, should be rewarded in a manner consistent with the aim of copyright to encourage creativity and innovation, and in the case of related rights, incentivizing investment. It is critical that AI systems function within the intellectual property systems that incentivize the development of high-quality input, regardless of whether the quality is a measure of vetting, structuring, or curation of the information used as input.

Investments in peer review, enhanced metadata, collection of disparate resources and the like may be secondary aspects of the application of AI, but they are critical to its success. In particular, those who invest in the creation, collection, or curation of data can and should be partners in ensuring the integrity and quality of AI systems and their output. For example, curators can help identify potential bias in the output based on the criteria that were used in creating collections in the first place.

3.2.1. STM best practice principles

- With respect to AI design and implementation, a great deal of the quality and integrity of AI systems can be addressed by adhering to principles of accountability and transparency. Sharing best practice principles will help publishers and other actors, including those involved in designing and implementing AI tools and services in scholarly communications, and also those writing code and creating algorithms. Additionally, persistent identifiers can help trace the source of code (see e.g., the standards established by the Software Heritage initiative).[10] Processes and data sets can include in their metadata how and when they were peer reviewed, tested and documented at each step; from planning, training and testing through to deployment.
- With respect to data input into AI systems, providing quality and integrity through various data validation and enhancement processes is at the core of the publishing process. To ensure that investments in data quality are protected and incentivized, it is important to formulate clear, commonly shared practices on content and data acquisition, use, and sharing (including licensing). In particular, an appropriate intellectual property framework will help incentivize the creation of high-quality IP that can be used as input data, as well as to protect pre-existing IP that might be used to create, train, calibrate, repair or improve AI systems. Copyrighted works will typically first have to be identified, selected, adapted, harmonized and normalized in order to be meaningfully deployed in any AI calibration. This requires policies to ensure that any works used are acknowledged and protected, with IP regimes that recognize their critical importance and rewards investment.
- With respect to output, it is very challenging to directly assess quality. Any limitations on the application of output should be clarified and made explicit to end users. A feedback option for users, offered when they are made aware of the use of AI in tools, processes and services, will help identify questionable or erroneous output so that AI system operators can take action. This includes a clear audit trail that links the output to the AI tool that created it. This will facilitate the resolution of errors or issues should these become apparent.

SPRINGER NATURE

Springer Nature is using a range of AI technologies to improve its data business. In a nutshell, this revolves around extracting facts from scientific texts and other sources, mine for and infer additional facts, and structure them into domain-specific ontologies and knowledge graphs for further application. Key technologies used for this purpose are NLP, ML, text classification, entity recognition and resolution, knowledge graphs and related semantic technologies. Use cases range from smart entity-based search & discovery in our database products, over assisted automation for content production, to producing data assets such as domain specific ontologies that power existing products and drive novel business cases.

[10] https://docs.softwareheritage.org/devel/swh-model/persistent-identifiers.html

3.3 / Privacy and Security

Privacy and security concerning the use of data in digital environments are becoming the subject of new legislation around the world. These new laws have obvious applications to AI systems. Publishers have a proven track record of commitment to the highest standards of quality and compliance with legal requirements. As fundamental rights, privacy and security should be at the forefront at all stages of AI architecture development, design and operation. In some territories, new legislation means that personal data protection is a legal requirement and has to be embedded throughout the development cycle for AI systems (e.g., the General Data Protection Regulation, GDPR,[11] in the EU).

AI adds a new level of complexity to any tool, process or service with respect to privacy and security. This is partly due to some of its opaque nature in certain applications, and partly because the outputs of AI-driven processes are often based on high volumes of data inputs (Big Data), making its provenance unclear. These inputs may contain personally identifiable information or other data where privacy considerations need to be taken into account. Even if the specific operations of AI algorithms and systems cannot be made fully 'open' and transparent, developers and operators of AI systems should be asked for transparency regarding the sources of underlying data used to train or drive those algorithms. Furthermore they should make clear that applicable laws regarding the collection and use of personal data have been followed.

Transparency about the provenance of data sources is at the heart of the publishing process. Key inputs are vetted, certified and validated in the publishing process, in secure data processing environments and in respecting authors' and contributors' privacy. Throughout their lifecycle, AI systems should respect and uphold privacy rights and data protection in a similar way. Just as important is ensuring the security of datasets used in training or operating such systems.

3.3.1. STM best practice principles

The following best practice principles and operational steps have been developed to ensure respect for privacy and data protection when designing, developing or using AI systems for data used and generated by the AI system throughout its lifecycle. If personal data is collected and used in AI systems, this should include maintaining privacy through appropriate data anonymization. The principles cover various important aspects involved in designing AI or engaging in data-sharing for AI purposes, or participating in AI system development, training, calibrating, learning or storing:

- Personal or sensitive data should at all times be handled in compliance with applicable privacy regulations (for example the General European Data Protection Regulation (GDPR).
- Sources of data should be made transparent.
- Adequate safeguards should be adopted to avoid corruption of data.
- Adequate measures should enable detection of tampering or manipulation of datasets.

[11] https://gdpr-info.eu

- Architecture that includes privacy safeguards, appropriate transparency and control over the use of data should be favored at all stages of development.
- The connection between data, and inferences drawn from that data by AI systems, should be sound and continuously assessed.
- Appropriate data and AI system security measures should be in place. This includes the identification of potential security vulnerabilities, and assurance of resilience to adversarial attacks.
- Security measures should account for unintended applications of AI systems, and potential abuse risks, with appropriate mitigation measures.
- Data, privacy and security impact assessments should accompany any project.

3.4 / Fairness

Bias can be an inherent problem in AI, not necessarily because of any malicious intent on the part of the designers or users of AI, but because of the very nature of the technology. In general terms, AI systems learn by identifying patterns in existing data. These historical patterns are subsequently used to make future predictions, recommendations, or decisions. But this carries the risk of replicating and amplifying historical bias. For example, AI will predict the performance of a person based on their behavior within a group with which they share certain traits and characteristics. While the list of "what is fair in AI" will never be exhaustive and may differ in different contexts, it should certainly avoid bias. This is why AI experts should not only review outputs for possible bias, but also constantly adjust systems to guard against the introduction of bias. This becomes ever more important in a world where an increasing number of decisions are made by or assisted by algorithms, in both the public and private sector.

This potential aspect of AI may be introduced or enhanced by the selection of training data with cultural or societal bias. Even where care is taken to identify and remove bias from datasets it is still possible for bias to emerge in the remaining data. For example, in a situation where age should not be considered relevant, this information may be removed from the data sample, but the algorithm may still combine several attributes to approximate a person's age (educational achievements, career development, spending habits, etc.). In other settings, bias may possibly enter input data from an orthogonal direction. For example, AI tools looking at fashion advertisements were found to have bias against people with disabilities, because they were trained to block depictions of medical devices to prevent misleading medical claims and were therefore blocking depictions of models who were using wheelchairs, prosthetics, and oxygen tanks even though those items were not being advertised.

STM publishers and organizations are a source of high-quality, vetted information that can form the basis for better, unbiased training and input data. Moreover, the application of metadata allows for a more careful selection of appropriate data. STM publishers are also excellent resources for informed review and evaluation of information through their investments in such processes as peer review.

As users of AI in tools, processes and services, publishers can help avoid bias. For example, without informed oversight, the application of AI in the review process could give authors from specific countries or institutions a negative recommendation based on the publication history of their compatriots or people from the same institutions. But there are other serious risks as well. AI tends to consolidate historical structures, which includes established scientific ideas and theories. The philosopher of science Thomas Kuhn has argued, however, that scientific breakthroughs are characterized by replacing paradigms with new ones (think of the heliocentric worldview of Kepler, Copernicus and Galileo, Darwin's theory of natural selection, and Einstein's theory of relativity). The risk of using a technology that looks at existing patterns to make predictions, recommendations or decisions, is that it suppresses the opportunity for new ideas to emerge, thereby stifling innovations and scientific breakthroughs.

Just as science is a continuous process, AI tools will equally need continuous review and refinement to ensure fairness and equity throughout their life cycle, from the data that is used to feed them, through their coding, to the outputs given and the application thereof. Continuous feedback loops are fundamental in ensuring that noise, bias or inaccuracy in input does not get replicated and amplified without the necessary checks and balances. STM publishers can contribute to this effort by developing standards, tools and processes that allow the evaluation of inputs and outputs for fairness.

3.4.1. STM best practice principles

- Identification of training or input data for internal purposes or to third parties should be done with care. Data must be carefully selected (e.g., using ontologies and metadata), reviewed and evaluated for potential sources of bias. If required, alternative data should be sought to correct for bias.
- The application of AI to tools, processes and services (both internal and external) should be carefully evaluated and reviewed in light of the inherent tendency of AI to reproduce and amplify existing patterns, potentially leading to bias, potential discrimination and the stifling of scientific innovation.
- In light of its potential dangers and risks, the application of AI and the data used to train and feed machines needs constant evaluation for potential bias and unfairness. In addition, there should be feedback mechanisms so that cases or concerns of bias can be reported to all stakeholders involved.

ELSEVIER

Elsevier is using AI technologies to enable its mission in delivering analytics to support researchers and healthcare professionals advance science and improve health outcomes for the benefit of society. The foundation of this work is the Entellect data integration platform that enables insights from data to drive effective innovation. The platform adheres to FAIR Principles, and facilitates access to clean, reusable data and metadata, enabling R&D scientists to optimize decision making. It enables better data governance and helps drive accurate AI/ML based discovery. A recent example of this work is the collaboration with the Pistoia Alliance and Mission Cure which used Elsevier data and 3rd party data to predict drug repurposing candidates for a rare disease.

3.5 / Sustainable Development

AI has often been heralded as a way to help solve long-term problems for humanity worldwide. Ending poverty, improving health and education, reducing inequality, tackling climate change, and spurring economic growth are all examples. There are various aspects to achieving human sustainability with the assistance of AI and each is an important factor in helping to realize the benefits that AI has to offer.

First is the ability of AI to contribute to projects of global scope that improve human lives and protect the natural environment for current and future generations. The human sustainability and ecological responsibility of AI systems should be further encouraged, and research should be fostered into AI solutions that address areas of global concern such as those identified in the United Nations Sustainable Development Goals.[12]

A good example is UN Goal 17, which focuses on the means of implementing the goals and the partnerships needed to deliver the technology, capacity and data required to measure and monitor progress. Goal 17.7 is to "promote the development, transfer, dissemination and diffusion of environmentally sound technologies to developing countries on favorable terms including on concessional and preferential terms, as mutually agreed." Considering that highly interdisciplinary approaches and the application of many data sources will be needed to achieve the goal, AI is especially promising because it traverses subject areas far better than practices confined to a single field.

Second is the potential of developing and using AI for both public and private organizations to ensure sustainability and foster efficiency in their own internal operations, as external environments change. Indeed, the deployment of AI systems in internal processes is crucial in empowering organizations to develop an awareness of inefficacies, costs, and choices – and how those can better materialize in a sustainable and long-term approach. To that effect, several processes could be automated and data concerning internal processes could be collected, organized, and analyzed to shed light on internal practices. AI solutions can be deployed to interpret huge datasets to extract actionable pieces of information from whole collections of research papers to support decision-making and ensure that interventions are evidence- and data-based. They can be used in the same way to identify appropriate solutions to organizational challenges.

Publishers play a role in providing and enhancing the quantity and quality of material optimized for use by AI systems. This is because their core expertise includes securely storing and organizing high-quality, structured information, tagging and enriching content and creating ontologies. These factors are all extremely valuable for successful AI – the more valuable the input, the more valuable the output of AI.

AI can be used to extract actionable knowledge and insights from publishers' large collections of scientific publications, creating decision-support tools for practitioners in medicine, agriculture, and other disciplines. A good example is CABI's PRISE (Pest Risk Information) service which combines information on plant-pest life cycles, earth

[12] https://sdgs.un.org/goals

observation (satellite) data and local knowledge to create highly practical alerting services for smallholder farmers in developing countries.[13]

STM member publishers are also at the forefront in using AI to streamline internal operations. They utilize machine learning to automate and rationalize their internal development and production processes, and also improve core areas of their operations such as assisting with the identification of peer reviewers, identifying, and combating plagiarism, recognizing fabricated data, and supporting the decision-making process behind the acceptance and rejection of manuscripts.

3.5.1. STM best practice principles

- While sustainable AI alone is not sufficient to ensure a successful, beneficial, and ethical outcome, it does have a crucial role to play in maximizing the benefits of AI to human health, welfare, and the natural world in which we live. Policy makers and research funders should create incentives for providers of key input data (e.g., in publications and databases) to engage in economically sustainable activities that support current and future AI development. This includes incentives to not only offer but also enhance the quantity and quality of material optimized for use by AI systems.
- A commitment to energy efficiency and, where possible, the use of renewable energy is another key component of sustainable AI tools. This commitment could include AI development that helps others realize their own energy goals by providing insights into the carbon emissions associated with the digital content, data and applications that run on an organization's servers or through the development of energy-efficient algorithms that reduce the number of power-hungry machines in AI processes.
- Sustainability should be a core principle of the AI tools themselves. The design and code associated with key AI solutions should be archived in secure, stable environments with sufficient documentation to enable developers other than its authors to understand and adjust the system as it evolves through time. For public organizations, this might mean storage in an open repository. For privately held organizations, this would mean availability to internal teams. Systems should also be designed so they can use required information from different, readily available sources to avoid overdependence on data that may be temporarily or permanently unavailable. To reduce duplication and waste, systems should be interoperable. The building blocks of these tools (including code, algorithms, etc.) should be based on community standards and best practice and where applicable should adhere to the FAIR principles (making data Findable, Accessible, Interoperable and Reusable).[14]

[13] https://www.cabi.org/projects/prise-a-pest-risk-information-service/

[14] https://www.nature.com/articles/sdata201618

About STM

At STM we support our members in their mission to advance research worldwide. Our over 140 members based in over 20 countries around the world collectively publish 66% of all journal articles and tens of thousands of monographs and reference works. As academic and professional publishers, learned societies, university presses, start-ups and established players we work together to serve society by developing standards and technology to ensure research is of high quality, trustworthy and easy to access. We promote the contribution that publishers make to innovation, openness and the sharing of knowledge and embrace change to support the growth and sustainability of the research ecosystem. As a common good, we provide data and analysis for all involved in the global activity of research.

www.stm-assoc.org

For more information about this publication, please contact
Joris van Rossum
Director of Research Integrity
at rossum@stm-assoc.org

STM Advancing trusted research