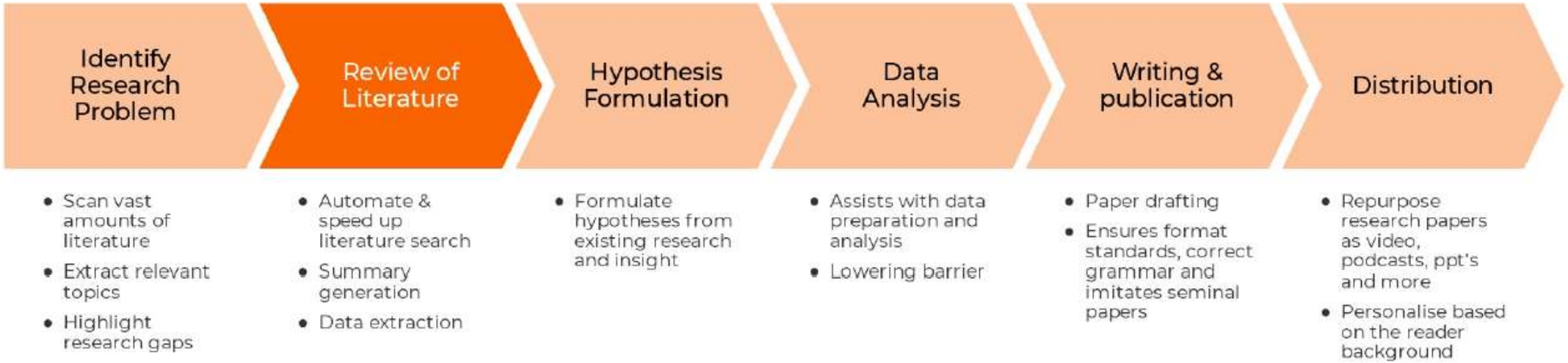
Universidade do Minho
Serviço de Documentação e Bibliotecas

Relevância e desafios da inteligência artificial na investigação e comunicação científica

Antónia Correia | 21 de maio de 2024

IA na investigação

- Rapidez de processamento de informação
- Geração automática de texto, imagens, …
- Disponibilidade 24/7
- Ajuda em trabalhos repetitivos, tradução e transcrição



Saikirai Chandra**, The impact of AI in research** Open Science Fair 2023

Algumas preocupações com a utilização da IA

- Alucinações
- Preconceitos
- “Atropelo” de direitos de autor
- Conhecimento disciplinar deficiente
- Utilização pouco ética de ferramentas de IA
- Qualidade e pertinência dos dados utilizados (dados “sujos”, dependência apenas de dados digitais, representatividade deficiente de alguns grupos)
- Qualidade e relevância dos resultados da investigação baseada em IA

Integridade na investigação

https://ukrio.org/research-integrity/what-is-research-integrity/

Comportamento não ético

Comportamento não ético - FFP

- **Fabricação**
- **Falsificação**
- **Plágio**

Fraude

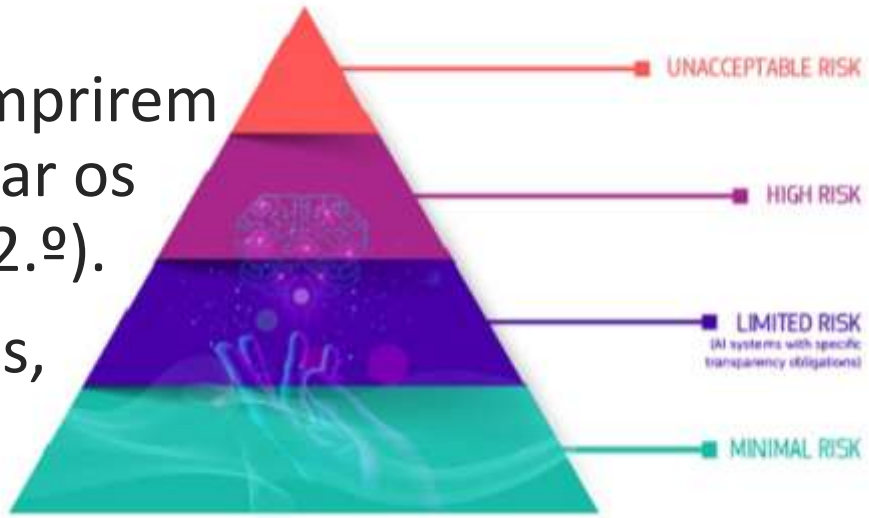
As práticas de investigação questionáveis incluem:

- **Representação incorrecta**
- **Inexatidão**
- **Preconceito**

"Ciência descuidada"

European Union AI Act

- **Sistemas de IA de risco inaceitável** são proibidos (Título II), por serem contrários aos valores da União.
- **Sistemas de IA de alto risco** são permitidos, embora sujeitos a obrigações (título III - risco elevado para a saúde e a segurança ou para os direitos fundamentais das pessoas singulares. Estão sujeitos a determinados requisitos obrigatórios e a uma avaliação de conformidade ex ante).
- **Sistemas de IA de risco limitado** são permitidos se cumprirem requisitos mínimos de transparência, tais como informar os utilizadores de que estão a interagir com a IA (artigo 52.º).
- **Sistemas de IA de baixo risco** podem ser desenvolvidos, produzidos e utilizados livremente.



Princípios HHH: Helpful, Honest, Harmless

Sistemas de IA que são úteis (Helpful):

- São treinados e aperfeiçoados tendo em conta as necessidades e os valores dos utilizadores
- Tentam claramente realizar a operação solicitada pelo utilizador ou sugerem uma abordagem alternativa quando a tarefa o exige
- Aumentam a produtividade, poupam tempo ou facilitam as tarefas dos utilizadores num determinado caso de utilização ou gama de casos de utilização
- São acessíveis a utilizadores com um amplo espetro de capacidades e conhecimentos

Sistemas IA que são honestos (Honest):

- Fornecem informações exatas quando podem e comunicam claramente aos utilizadores quando não podem produzir um resultado exato
- Expressam a incerteza e a razão por detrás dela
- São desenvolvidos e funcionam de forma transparente para que os utilizadores possam compreender como funcionam e confiar no que produzem

Quadro ético para a utilização da IA generativa

Princípios HHH: Helpful, Honest, Harmless

Sistemas de IA que são inofensivos (Harmless):

- Não obedecem quando lhes é pedido que executem uma tarefa perigosa
- São treinados em estruturas que atenuam de forma transparente e ativa os preconceitos
- Não discriminam nem demonstram preconceitos explícita ou implicitamente
- Comunicam de forma sensível quando se envolvem com um utilizador num tópico sensível

A melhor forma de reforçar a utilidade, honestidade e inocuidade de um modelo é através do **feedback humano**. A aprendizagem por reforço a partir de feedback humano (RLHF) e o ajuste fino supervisionado (SFT) são duas técnicas cada vez mais populares para alinhar modelos com a HHH em mente.

https://www.invisible.co/blog/helpful-honest-harmless-ai

Utilização responsável da IA na investigação

2.1. RECOMMENDATIONS FOR RESEARCHERS

For generative AI to be used in a responsible manner, researchers should:

1. **Remain ultimately responsible for scientific output.**
   - Researchers are accountable for the integrity of the content[13] generated by or with the support of AI tools.
   - Researchers maintain a critical approach to using the output produced by generative AI and are aware of the tools' limitations, such as bias, hallucinations[14] and inaccuracies.
   - AI systems are neither authors nor co-authors. Authorship implies agency and responsibility, so it lies with human researchers.
   - Researchers do not use fabricated material created by generative AI in the scientific process, for example falsifying, altering or manipulating original research data.
2. **Use generative AI transparently.**
   - Researchers, to be transparent, detail which generative AI tools have been used substantially[15] in their research processes. Reference to the tool could include the name, version, date, etc. and how it was used and affected the research process. If relevant, researchers make the input (prompts) and output available, in line with open science principles.
   - Researchers take into account the stochastic (random) nature of generative AI tools, which is the tendency to produce different output from the same input. Researchers aim for reproducibility and robustness in their results and conclusions. They disclose or discuss the limitations of generative AI tools used, including possible biases in the generated content, as well as possible mitigation measures.
3. **Pay particular attention to issues related to privacy, confidentiality and intellectual property rights when sharing sensitive or protected information with AI tools.**
   - Researchers remain mindful that generated or uploaded input (text, data, prompts, images, etc.) could be used for other purposes, such as the training of AI models. Therefore, they protect unpublished or sensitive work (such as their own or others' unpublished work) by taking care not to upload it into an online AI system unless there are assurances that the data will not be re-used, e.g., to train future language models or to the untraceable and unverifiable reuse of data.

[Link]

- Researchers take care not to provide third parties' personal data to online generative AI systems unless the data subject (individual) has given them their consent and researchers have a clear goal for which the personal data are to be used so compliance with EU data protection rules[16] is ensured[17].
- Researchers understand the technical and ethical implications regarding privacy, confidentiality and intellectual property rights. They check, for example, the privacy options of the tools, who is managing the tool (public or private institutions, companies, etc.), where the tool is running and implications for any information uploaded. This could range from closed environments, hosting on a third-party infrastructure with guaranteed privacy, to open internet-accessible platforms.

4. **When using generative AI, respect applicable national, EU and international legislation, as in their regular research activities.** In particular, the output produced by generative AI can be especially sensitive in relation to the protection of intellectual property rights and personal data.
   - Researchers pay attention to the potential for plagiarism (text, code, images, etc.) when using outputs from generative AI. Researchers respect others' authorship and cite their work where appropriate. The output of a generative AI (such a large language model) may be based on someone else's results and require proper recognition and citation[18].
   - The output produced by generative AI can contain personal data. If this becomes apparent, researchers are responsible for handling any personal data output responsibly and appropriately, and EU data protection rules are to be followed.
5. **Continuously learn how to use generative AI tools properly to maximise their benefits, including by undertaking training.**
   - Generative AI tools are evolving quickly, and new ways to use them are regularly discovered. Researchers stay up to date on the best practices and share them with colleagues and other stakeholders.
6. **Refrain from using generative AI tools substantially[19] in sensitive activities that could impact other researchers or organisations (for example peer review, evaluation of research proposals, etc).**
   - Avoiding the use of generative AI tools eliminates the potential risks of unfair treatment or assessment that may arise from these tools' limitations (such as hallucinations and bias).
   - Moreover, this will safeguard the original unpublished work of fellow researchers from potential exposure or inclusion in an AI model (under the conditions detailed above in the recommendation for researchers #3).

- Humanos são responsáveis pelos resultados de investigação;
- O uso de ferramentas deve ser reportado
- Informação sensível não deve ser importada para ferramentas de IA
- Conhecer a legislação nacional, europeia e internacional
- Aprender a usar a IA
- Não usar IA para tarefas com impaco direto em terceiros, ex. revisão por pares e avaliação de propostas de investigação

Autoria e ferramentas IA

- As ferramentas de IA, como o Chat GPT, não podem ser listadas como autores de documentos, uma vez que não podem assumir a responsabilidade pelo trabalho apresentado
- Os autores devem revelar explicitamente quando utilizam estas ferramentas

Home

Authorship and AI tools

COPE position statement

The use of artificial intelligence (AI) tools such as ChatGPT or Large Language Models in research publications is expanding rapidly. COPE joins organisations, such as WAME and the JAMA Network among others, to state that AI tools cannot be listed as an author of a paper.

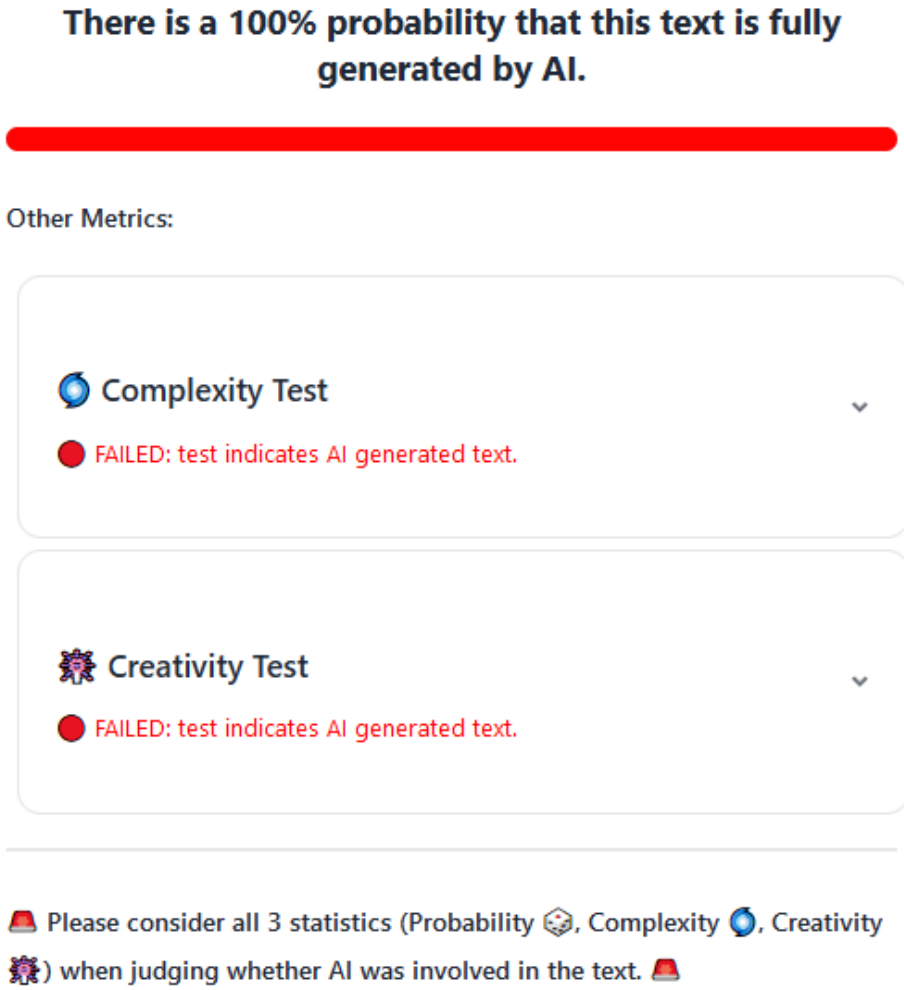
AI tools cannot meet the requirements for authorship as they cannot take responsibility for the submitted work. As non-legal entities, they cannot assert the presence or absence of conflicts of interest nor manage copyright and license agreements.

Authors who use AI tools in the writing of a manuscript, production of images or graphical elements of the paper, or in the collection and analysis of data, must be transparent in disclosing in the Materials and Methods (or similar section) of the paper how the AI tool was used and which tool was used. Authors are fully responsible for the content of their manuscript, even those parts produced by an AI tool, and are thus liable for any breach of publication ethics.

https://publicationethics.org/cope-position-statements/ai-author

Detecting text generated by chatbots

- GPT4 Detector .ai
- AI Text Classifier
- GPTZero
- Winston AI



Technology

Plagiarism tool gets a ChatGPT detector – some schools don't want it

Popular plagiarism detection software used by many schools and universities worldwide is set to get an AI-detecting component in the wake of the release of ChatGPT

By Jeremy Hsu

3 April 2023

TECH IN YOUR LIFE

We tested a new ChatGPT-detector for teachers. It flagged an innocent student.

Five high school students helped our tech columnist test a ChatGPT detector coming from Turnitin to 2.1 million teachers. It missed enough to get someone in trouble.

Analysis by Geoffrey A. Fowler

Columnist | + Follow

Updated April 3, 2023 at 9:47 a.m. EDT | Published April 3, 2023 at 6:00 a.m. EDT

FUTURE

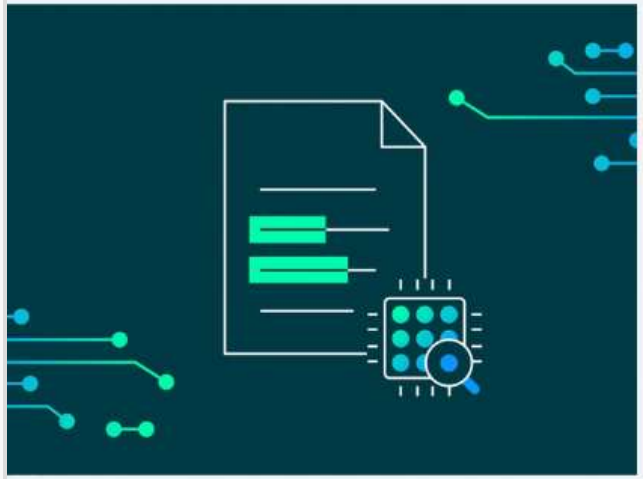
How to Prove You Didn't Use ChatGPT: One Simple Trick to Avoid ChatGPT Plagiarism Accusations

By Amy D

Posted on May 22, 2023

turnitin

Products Solutions Resources Support Contact Sales



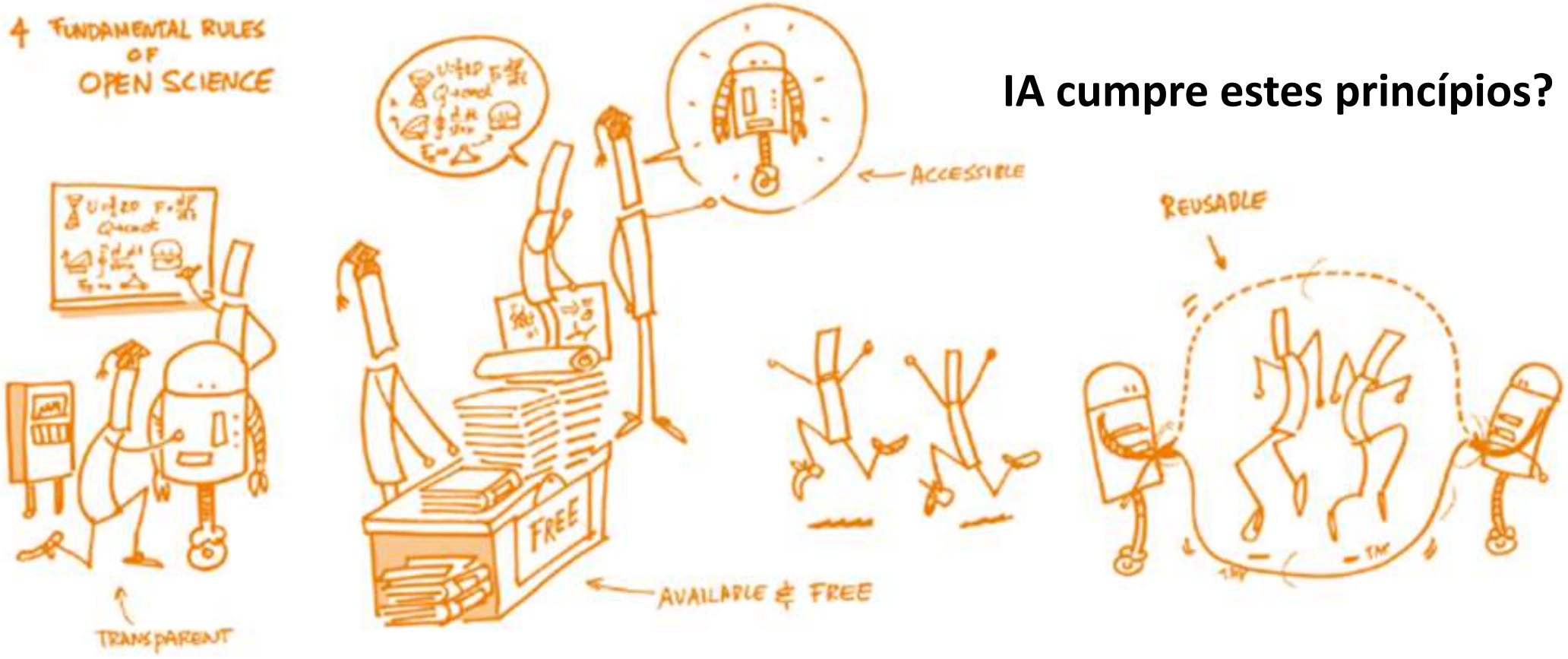
Blog · 3 min read · Tuesday 4 April 2023

The launch of Turnitin's AI writing detector and the road ahead

"Today, we are pleased to announce the launch of our AI writing detection capabilities in Turnitin Feedback Studio (TFS), TFS with Originality, Turnitin Originality, Turnitin Similarity, Simcheck, Originality Check, and Originality Check+. The detector will support over 2.1 million educators and more than 10,700 institutions, reaching more than 62 million students. It is a milestone that represents an incredible commitment from the Turnitin team as well as our customers."

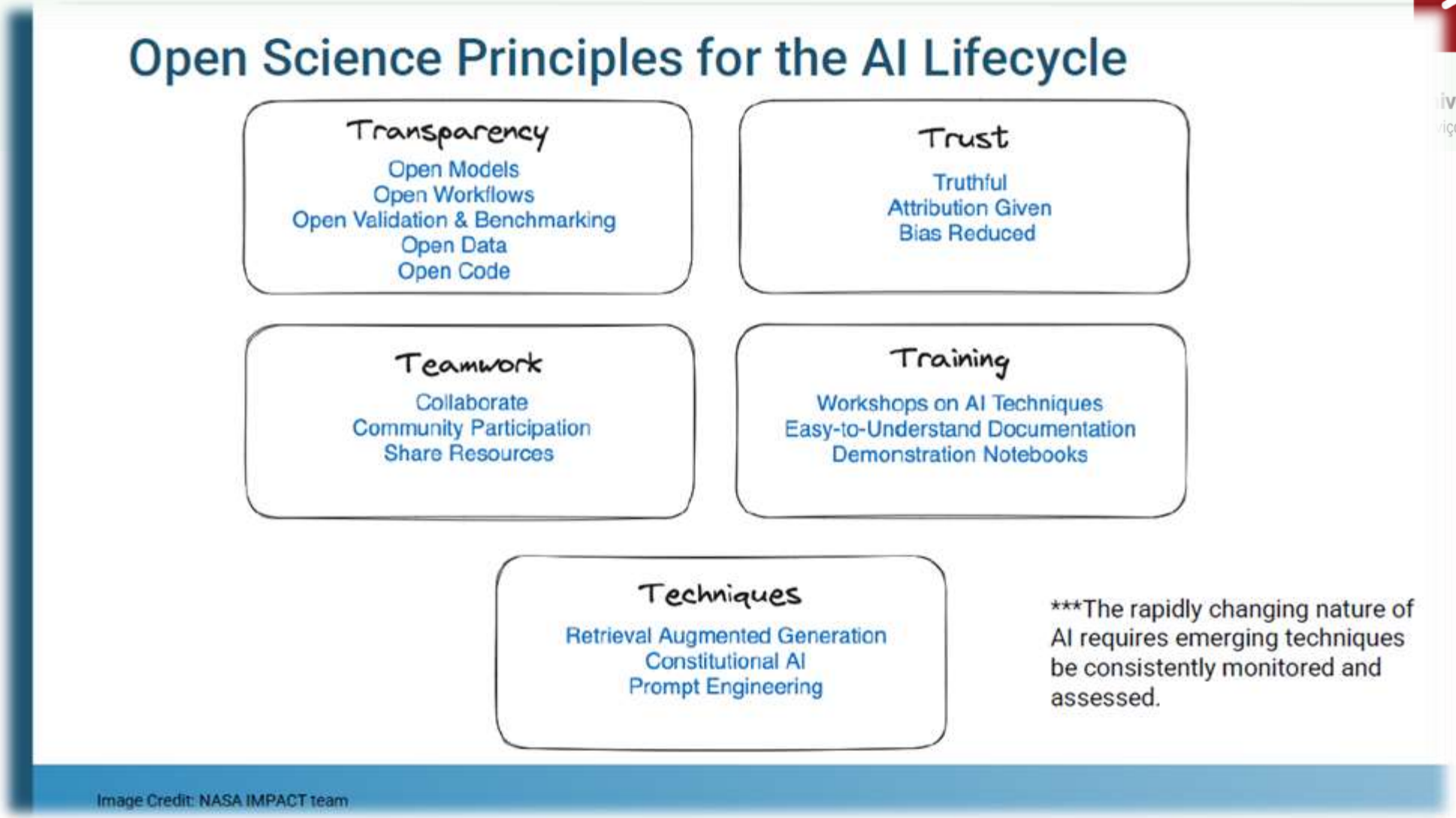
https://www.turnitin.com/blog/the-launch-of-turnitins-ai-writing-detector-and-the-road-ahead

Princípios fundamentais da Ciência Aberta



IA cumpre estes princípios?

Como pode a CA contribuir para a melhoria dos sistemas e ferramentas de IA?



Kaylin Bugbee. Architecting the Future: NASA's Use of Large Language Models to Enable Open Science, **Open Science Fair 2023**

Relevância para os Bibliotecários

Necessidades

- Estar informado e tentar compreender o funcionamento e as limitações das ferramentas de IA
- Testar, demonstrar, analisar
- IA também utiliza dados abertos para treinar algoritmos – investir na curadoria de dados
- Contexto RRI - privacidade, preconceitos, plágio, questões éticas
- Conhecer as melhores práticas/recomendações
- Destacar os desafios e as questões em aberto

Desafios

- Âmbito alargado e pouco claro (o que é a IA, o que não é a IA)
- É necessário conhecimento/compreensão técnica
- Perspectivas diversas e percepções dos formandos
- Explicar aos investigadores que a qualidade dos dados é importante
- Sensacionalismo nos media
- Percepções individuais

Universidade do Minho
Serviço de Documentação e Bibliotecas

Antónia Correia
antonia.correia@usdb.uminho.pt