

## **PERSPETIVAS SOBRE COPYRIGHT E PLÁGIO, NO CONTEXTO DE GENERATIVE ARTIFICIAL INTELLIGENCE**

### **Perspectives on copyright and plagiarism, in the context of Generative Artificial Intelligence**

**Artur Marques**

Instituto Politécnico de Santarém - Escola Superior de Gestão e Tecnologia, Portugal

CIAC – Polo de Literacia Digital e Inclusão Social, Portugal

[artur.marques@esg.ipsantarem.pt](mailto:artur.marques@esg.ipsantarem.pt)

**Filipe Madeira**

Instituto Politécnico de Santarém - Escola Superior de Gestão e Tecnologia, Portugal

CIAC – Polo de Literacia Digital e Inclusão Social, Portugal

[filipe.madeira@esg.ipsantarem.pt](mailto:filipe.madeira@esg.ipsantarem.pt) | ORCID: 0000-0002-2227-7006 | Ciência ID:DE1F-7FEE-FBA5

## **RESUMO**

Embora a Inteligência Artificial para a geração de conteúdos, ou Generative AI (GAI), tenha décadas de maturidade, é na década de 2020 que atinge uma massa crítica tal, que não é fácil identificar sectores de atividade em que não tenha aplicações e implicações.

Este artigo apresenta duas primeiras perspetivas, aparentemente contrárias às opiniões mais populares - conforme aferido por análise de sentimento em rede social -, sobre dois problemas que ganharam visibilidade com a massificação de sistemas de GAI e de Machine Learning (ML):

- O problema da atribuição de copyright para trabalhos por estes sistemas;
- Aplicações para controlo de plágio, por ML.

O sentimento presente parece ser que:

- Trabalhos por sistemas de GAI não podem ser protegidos por copyright, porque tal atribuição só está disponível para autores humanos, facto corroborado em casos como o do fotógrafo David Slater e o auto-retrato de uma simio com o seu equipamento, que levou o United States Copyright Office a recordar que “trabalhos por ... entidades não-humanas não são elegíveis para protecção”.
- A deteção de plágio atingiu níveis de disponibilidade e aceitação tais, que ferramentas baseadas em AI, para Learning Management Systems (LMSs), são agora confiadas por muitos educadores.

A nossa perspetiva está desalinhada destas opiniões dominantes, podendo ser resumida assim:

- Na realidade, há protecção de copyright para qualquer trabalho humano com um mínimo de criatividade atribuível, pelo que ferramentas de GAI que assistam o processo são legítimas, pois o simples esforço de identifica-las, seleccioná-las e aplica-las em partes da criação deverá satisfazer os requisitos, com precedentes em casos como a protecção concedida a diretórios de conteúdos, como listas telefónicas: no caso “Feist Publications, Inc. vs. Rural

Telephone Service Co., Inc.”, escreve-se explicitamente que “... seleção e arranjo ... constituem esforço criativo mínimo”.

- Quanto à deteção de “plágio” por ferramentas de AI, na realidade a expressão comumente refere-se à deteção de similaridades, e não ao radical do termo, que é o não reconhecimento de trabalho alheio. Assim, quanto à mensuração de quanto é que um trabalho é copiado; em particular, se trabalhos de texto são produzidos por sistemas de GAI, no limite essa deteção é impossível e tais soluções são não confiáveis.

É quase paradoxal que sejam técnicas de GANs (Generative Adversarial Networks) a, simultaneamente, justificar a confiança em soluções para este propósito, e também a justificarem a perspectiva de que não há solução, a prazo.

Uma GAN é, na sua essência, a oposição entre um sistema gerador e um sistema avaliador ou discriminador. O avaliador tem “vasto” conhecimento de determinados trabalhos, como textos ou fotografias verdadeiras; o gerador começa por produzir lixo aleatório, que o discriminador detetará facilmente como uma falsidade. Ao longo da oposição, ambas as partes vão melhorando, a gerar e a detetar falsidades, mas a partir de certa iteração, o expectável é que o gerador produza conteúdos indistinguíveis da realidade. Assim, a mesma abordagem que constitui solução para a deteção, é solução para derrotar-se a si própria.

**Palavras-chave:** Direitos de autor, Plagiarismo, Aprendizagem automática, GANs, Análise de sentimento

## ABSTRACT

Although Artificial Intelligence for Content Generation, or Generative AI (GAI), is decades old, it is in the 2020s that it reaches such critical mass, that it is not easy to identify activity sectors immune to its applications and implications.

This article presents two perspectives, somewhat contrarian to popular opinion - as measured by social network sentiment analysis - on two problems that gained visibility with the massification of GAI and Machine Learning (ML) systems:

- The problem of copyright attribution to works based on these systems;
- ML-based applications for plagiarism control.

The current perception seems to be that:

- GAI-based creations cannot be protected by Copyright, because such protection is only available to humans, as seen in cases such as the photographer David Slater and the self-portrait of a macaque using his equipment, in which the United States Copyright Office recalled that “works by ... non-human entities are not eligible for copyright protection”.
- Plagiarism detection has reached such levels of availability and acceptance, namely in the form of AI-based tools for Learning Management Systems (LMSs), that they are now trusted by many educators.

Our perspective is misaligned from these opinions, and can be summed up as follows:

- In fact, there is Copyright protection for any human-attributable work with a minimum of creative effort, so GAI-based tools that assist the process, are legitimate. The mere effort to identify, select and apply them should meet the requirements, with precedent in cases such as “Feist Publications, Inc. vs. Rural Telephone Service Co., Inc. ”, in which protection was granted to content directories, such as phone numbers, where it was considered that the “... selection and arrangement ... constitute minimal creative effort ”.

- Regarding the detection of “plagiarism” by AI tools, in fact the expression commonly refers to the detection of similarities, not to the radical meaning of the term, which is the non-recognition of the work of others. Regarding assessing how much of a work is copied, in particular if some text is the output of an AI system, in the limit this detection is impossible and the available solutions are not reliable.

Almost paradoxically, it is the same GANS (Generative Adversarial Networks) that eventually support the confidence in solutions for this purpose, that also justify the perspective that there is no solution, on the long run. A GAN, is the opposition between a generator and an evaluator or discriminator system. The evaluator has “vast” knowledge of certain works, such as texts or real photographs; the generator begins by creating random outputs, which the discriminator will easily detect as a fake. As the process runs, both parts get better, but from a certain iteration, the generator is expected to produce content indistinguishable of reality. Thus, the very same approach that offers a solution to detection, is a solution to its defeat.

**Keywords:** Copyright, Plagiarism, Machine Learning, GANs, Sentiment analysis