Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

How original is artificial intelligence?

Consultation response of the European Writers' Council, registered observer at WIPO, to WIPO/IP/AI/2/GE/20/1 - WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI). To be submitted until 14 February 2020.

Brussels, 14.02.2020

We thank you for the opportunity to comment on the topic of Artificial Intelligence and Intellectual Property during the ongoing conversation. In the following, we will not only answer questions within Issues 6, 7, 9 and 10, raised in document WIPO/IP/AI/2/GE/20/1, but we will also take the chance of making comments where we wish to make corrections to the theses set out.

The European Writers' Council represents the interests of 150,000 authors in the book and text sector from 41 organisations in Europe and non-EU countries who write and publish in 31 languages and in all genres.

Preamble: An overview of the concept of "artificial intelligence" and the current use of machine-learned natural language understanding (NLU), natural language processing (NLP) and natural language generation (NLG) in the book and text sector

The term "artificial intelligence" is controversial. The subarea of computer science called AI has nothing to do with intelligence, which we humans or even animals recognize, and with which competencies such as consciousness, decisionmaking, cognitive abilities and abstract, multidimensional processes are associated. Mostly it is about pattern recognition and curation of large amounts of data, it's mathematics, probability calculation.

A distinction is made between «strong» and «weak» AI. Strong AI has the goal of making the mechanical reproduction of psychological processes such as thinking, learning or problem solving possible. There is currently no application that can handle this.
Weak AI, on the other hand, is strictly limited to certain sub-sectors. These generators of weak AI already exist - but forms of strong AI do not. Nevertheless, many people, when they read about AI, especially when it comes to produce or translating text works, have the idea of «strong AI» in their minds. Just as the theses and questions of your paper suggest, too.

We would like to point out that the development of strong AI in the text area is reaching the limits of the simulation of psychological, emotional and cognitive human functioning will. Although there are already working speech recognition and translation AI, they have nothing to do with scientific or fictional manuscript work.

In some cases, text-generating systems are referred to as «weak artificial intelligence», which - contrary to all intelligence as we classify it - are purely rule-based, without independent, conscious or

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

original creative creation processes. Like the «automatic filler» that you know from your mobile phones. This «Markov chain» calculates the probability of the next word, and was developed in the 1960s. To this day, it still uses a limited vocabulary. Even more advanced RNN (Recurrent Neural Networks) systems, which remember the context of a very short text to get to the most likely next word, fail because of longer sentences and become confused.

The text generation of the «Natural Language Generation» works through a template-based approach[1]: While a grammar-based approach would try to simulate the human intellect to write texts of any genre independently, the template approach uses pre-formulated text parts that are automatically reassembled by a human story plot (!): a word puzzler. He cannot form puzzle pieces himself or even think up the picture automatically, he needs guard rails and access to a collection of existing text modules created by human authors. The current vocabulary of NLG is restricted.

Currently, these systems are gradually being replaced or supplemented by deep machine learning systems. In the case of text generation, Machine Learning and Natural Language Understanding supports a user, for example, through linguistic analysis during text creation, controls word repetition and other parameters like grammar or spelling: here we are talking about supporting work of a program that also may summarize long texts, information, minutes of conversations, etc.
This is sometimes used in the publishing houses to produce information for cover graphics, to summarize a novel in this way and to extract basics for blurb texts or advertising.
Again, these are not independent works.

In the book sector, weak algorithmic programs and template text generators are used, based on pattern recognition and the use and evaluation of existing text modules. Examples:
Since three years, the weak AI «QualiFiction» has been testing manuscripts for their potential for success by comparing them with existing bestsellers. The bestseller formula has unfortunately still not been found.
Text generators and robot journalism, such as the weak template AI of the company uNaice, are used to create short texts of products on online shop pages. In China, bots - programs that crawl through the Internet and collect material - compile text collections from Wikipedia entries.[2] Chat bots - i.e. text-based dialogue system - conduct conversations with customers on websites and use a repertoire of standing answers. Self-learning chat bots react to trigger words in social media and, depending on the programming, emit either enthusiasm or hate tweets[3], pretending to be a human account.
They cannot decide what is good or bad, true or false. Weak NLG curates, recognizes patterns and assembles text modules on the basis of probability.

Translation programs learn faster[4] - but their word-for-word translation takes no account of allusions, metaphors, irony, the nuances of synonyms or sentence rhythm. So DeepL can translate for example shorter texts in one of big nine world languages bases, but not without an editorial and linguistic like narrative intensive revision, around insinuations, bon mots or slang adequately into the other culture area transfer. Even for minority languages, there will most likely not be any neural

[1] https://medium.com/sciforce/a-comprehensive-guide-to-natural-language-generation-dd63a4b6e548
[2] https://www.oracle.com/solutions/chatbots/what-is-a-chatbot/
[3] https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist
[4] https://medium.com/sciforce/neural-machine-translation-1381b25c9574

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

network machine translation, because the transfer data is too small to learn from them. In the long run this would mean, we are facing a loss of minority languages.

Some Member State governments already use automatic translators like DeeplL for legislative texts. There is a high risk here that, due to mechanical shortcomings, different interpretations of a law are in circulation and lead to disastrous consequences.

Accordingly, we are astonished that some publishers sell automatically translated books, saving themselves fees while claiming copyright on the translations for themselves. It is also unclear how automatic translations of news pages are linked to the neighbouring copyright law to press oublishers in order to remunerate further use of journalistic texts accordingly.

There will be even greater legal loopholes and literary accidents (pardon) in the future, and so we welcome WIPO's move towards coordinated conversation.

With GPT2, the artificial intelligence of OpenAI / Microsoft[5], it should be possible, according to self-disclosure, which is of course also self marketing, that an AI, which has NLG and NLU as well as self-referential capabilities, writes a whole novel one day (after human input of parameters, of course). If one feeds GPT2 today with the first sentences of George Orwell's novel «1984», which read: «It was a radiantly cold April day, and the clocks were striking thirteen», - the probability calculator adds: «I was in my car on the way to my new job in Seattle. ... I just imagined what the day would be like. A hundred years from now.»

What could this day look like when we want to talk and decide about authors' rights, personal rights, copyright and machine-learned text generators and natural language processing today?

Issue 6: Authorship and Ownership

WIPO:
12. «AI applications are capable of producing literary and artistic works autonomously.»

> **EWC comment**
>
> We do not agree with this thesis.
>
> Machine-generated texts do not «create autonomously» a «literary or artistic work», since this is characterized by an original human creation.
>
> Machine-generated texts do not produce anything new, nothing of their own; they refer exclusively to existing originals. Curated or randomly generated text forms also refer to and draw exclusively on and from existing self-created and individually created, copyrighted, and - this is another aspect - legally compliant works. Legal conformity with regard to, for example, personal rights or citation. Here, the aspect of responsibility must also be considered, which human authors pay attention to, which machine-generated text generators cannot provide.

---

[5] https://msturing.org

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

And so machine-generated text forms add no additional knowledge, no original and self-creative, self-experienced, authentic content to the literary and text world, uniquely coloured by individuality in the present.

Therefore, the results of text robots are by no means to be equated with literary or artistic works. This is also a crucial point for the future consideration of copyright and artificial intelligence.

Here we ask for a clear correction of this inaccurate basic thesis that machine-generated text forms are «autonomously designed literary or artistic works», because the opposite is the case: None of them are autonomous, but everything is copied; not a single observation has been created in an autonomous way.

WIPO:
«This capacity raises major policy questions for the copyright system, which has always been intimately associated with the human creative spirit and with respect and reward for, and the encouragement of, the expression of human creativity. The policy positions adopted in relation to the attribution of copyright to AI-generated works will go to the heart of the social purpose for which the copyright system exists.»

EWC comment

We take the opportunity of noting that creativity and the human spirit are only two components for creating literary and non-literary textual works in an original, new way and with an inimitable level of creation of their own. We clearly see in the two subordinate clauses of your draft paper an all too common underestimation of the effort required to create a work. And at the same time an overestimation of NLG and its currently exclusively weak forms.

We would like to outline these human achievements briefly to make clear that it is less about a religious war where man and machine face each other and are compared. This comparison is, if one knows the activities and challenges necessary to create a work, already impossible - precisely where machines with their limited sub-areas reach their limits - for example, where an NLG can write but cannot read (itself).

- Writing is an act of consciousness as well as subconsciousness, psyche, cognitive and emotional abilities, constantly updated knowledge and the self-reflective self. We call this personality, through whose culturally individual resonance space, through their own impressions and their processing, and through whose home of values formed from the course of life, the work and its expression is filtered. In this way, original works are created that are unique because the person who created them is unique. A machine has no consciousness, no curriculum vitae, no original collection of uniquer experience.

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

- Writing is an act that requires, in addition to linguistic and dramaturgical craft, to know forgotten and new metaphors, to take up and use allusions from one's own culture and from the analogue and digital dynamic present, and to devote oneself to the nuances of irony, sarcasm and exaggeration. Their understanding is based on conscious experience. A machine does not recognize irony.

- Writing requires self-authorisation. Whether vanity, need, urge, desire to impart knowledge; out of discrimination or self-defence against a dictatorial regime. Raising the voice in writing, the reason for which a work is created, contributes to the colour, the weighting, the content, the reading experience of a work. A machine has no own drive.

- Writing quality grows with the doing itself, with the aging of the personality, its change. It grows on itself, since all learning processes - training, attempts to imitate idols - lead to an own approach. Machines are plagiarizing. People learn and create something new from what they have learned.

- Writing is subject to a canon of values and norms of social coexistence. This implies truthfulness, accuracy, facts and reliable and balanced information, especially in the field of educational and academic books. A machine does not know what is right, and does not even know whether what it writes is right or wrong, true or false. It will never be able to make this decision independently. This can be used for abuse.

- In copyright and authors' rights laws, the copyright owner and therefore author and creator, is responsible for what is written and published and how. A machine cannot assume any duty of responsibility, and has also forfeited the right of authorship.

WIPO:

«If AI-generated works were excluded from eligibility for copyright protection, the copyright system would be seen as an instrument for encouraging and favoring the dignity of human creativity over machine creativity. If copyright protection were accorded to AI-generated works, the copyright system would tend to be seen as an instrument favoring the availability for the consumer of the largest number of creative works and of placing an equal value on human and machine creativity.»

EWC comment

We would like to point out that «creativity» is not the basis of machine-generated text works. Among other things, creativity is the ability to form and understand new, never-heard sentences, which is linked to linguistic competence. It is the ability to create something that is new or original and at the same time useful or usable.

In connection with works of art, it is also an achievement that is extraordinarily shaped by personality.

The automated, probabilistically calculated stringing together of letters of the weak AI is rather about: forms of curation, pattern recognition and pattern repetition, reference and reuse, plagiarism, forms of quotation, text module combinations, etc.

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

They do not create any work of their own without human guidance.
Developers of AI who produce texts see it as a long-term challenge to ever develop an AI that writes even remotely, or even: out of their own empowerment and will.

Specifically,
(i) Should copyright be attributed to original literary and artistic works that are autonomously generated by AI or should a human creator be required?

EWC comment:
No. The question is also a suggestive question and based on misguided assumptions. Weak AI generate no original literary or artistic work, and strong AI will not either.

(ii) In the event copyright can be attributed to AI-generated works, in whom should the copyright vest? Should consideration be given to according a legal personality to an AI application where it creates original works autonomously, so that the copyright would vest in the personality and the personality could be governed and sold in a manner similar to a corporation?

EWC comment:
Artificial intelligence does not produce an original work independently but refers exclusively to existing works. These works, in turn, which were used for so-called deep learning, are usually protected by copyright. This includes, in addition to the obligation to name the authors, and depending on national jurisdiction, an obligation to pay a fee for use.

So the question is irrelevant, because where is no independent, original creation, there is no copyright.

(iii) Should a separate sui generis system of protection (for example, one offering a reduced term of protection and other limitations, or one treating AI-generated works as performances) be envisaged for original literary and artistic works autonomously generated by AI?

EWC comment:
No. Artificial intelligence does not produce an original work independently.
Alternative or sub-protection systems would mean putting the creators of the works from which AI made use of them in a worse position than the machine itself, if the latter was provided with its own protection system, but at the same time insufficiently protecting the creators of the reference works from access by the machine.

In this respect, national jurisprudence must clearly be on the side of human authors, granting them, inter alia, opt-out procedures and legal remuneration, and also participation in any monetary transactions made with an AI that learns, copies, quotes and re-uses from them.

Issue 7: Infringement and Exceptions

13. An AI application can produce creative works by learning from data with AI techniques such as machine learning. The data used for training the AI application may represent creative works that are subject to copyright (see also Issue 10). A number of issues arise in this regard, specifically,

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

> **EWC comment:**
> Again we point out that a weak AI does not produce any «creative works».

(i) Should the use of the data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright? If not, should an explicit exception be made under copyright law or other relevant laws for the use of such data to train AI applications?

> **EWC comment:**
> We do not recommend further exceptions or limitations in national copyright frameworks on the use of copyrighted works or works which underly the authors' rights.
> Authors should be able to voluntarily allow or deny their works to be used, in a self-determined manner and with a legally established right to remuneration – but not under an exception. This voluntarily opt-in applies with regard to deep learning by text robots from original text works.

(ii) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, what would be the impact on the development of AI and on the free flow of data to improve innovation in AI?

> **EWC comment:**
> This is again a suggestive question, because the sub-meaning behind this accusation is, that authors' rights and copyright laws hinders innovation.
>
> We would like to strongly object to this meme from a well-known political camp and stakeholder area.
>
> Developments and the discovery of new possibilities within institutions that want to be innovative also have every opportunity to draw on the services of authors - under fair and clear conditions. These are opt-in, remuneration and: transparency.
> Under no circumstances should copyright infringements, which is also a violation of investments, be considered less than innovation intentions.
> This does not have to be done at the expense of authors in order to save these companies and institutions expenses and benevolent behaviour.
> Because then the artificial intelligence of the future will also have little ethical backbone: because that is what it is all about. About ethics and social standards for the behavior and reactions of automated, machine-controlled applications.
> In this case, the developers should already take great care in their own behavior.

(iii) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an exception be made for at least certain acts for limited purposes, such as the use in non-commercial user-generated works or the use for research?

> **EWC comment:**
> No. We do not recommend any further exceptions or limitations. Authors already accept many

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

exceptions and limitations, but at the same time have little protection and enforcement power. They are in the weakest position in the book chain, although the whole industry is build on them. They should not fund research and government education challenges. If AI is so highly valued, then those who make it possible with their texts and books should also be highly valued.

(iv) If the use of the data subsisting of copyright works without authorization for machine learning is considered to constitute an infringement of copyright, how would existing exceptions for text and data mining interact with such infringement?

> **EWC comment:**
> The conclusion that this is a copyright infringement is understandable. The EWC also strongly recommends that in implementing the articles of the DSM Directive, authors should be able to opt-in voluntarily and in any case be legally obliged to pay for the use of text and data mining.

(v) Would any policy intervention be necessary to facilitate licensing if the unauthorized use of data subsisting in copyright works for machine learning were to be considered an infringement of copyright?

> **EWC comment:**
> Again a suggestive question. Where collecting societies exist, licences can already be administered through them. Where there are no collecting societies, this is a good reason to establish them with political support, and at the same time to establish them legally and socially for the remuneration of authors for all uses of their works.

(vi) How would the unauthorized use of data subsisting in copyright works for machine learning be detected and enforced, in particular when a large number of copyright works are created by AI?

> **EWC comment:**
> Here, two things are related to each other that are by no means equivalent! There are no creative, independent works of an AI that are subject to copyright protection. Likewise, no originator is assigned to their non-creative, non-independent text quantities that do not correspond to the required threshold of originality pursuant.
>
> On the aspect of the deletion of unjustifiably used original works by authors for machine learning: Here a database, a HUB, affiliated for example with the EUIPO, can be useful for Europe. There, only works could be made available whose authors have voluntarily agreed (opt-in) to make their works available for machine learning for a limited period of time
>
> in return for fair remuneration and transparent information on how they are used.
> The rules for such a hub should never be established without representatives of major authors' federations. Nor should this be decided by publishers, but by those on whose work the ability to learn is trained. Authors want to have the freedom to say no to machines that are supposed to replace us one day – in the eyes of some stakeholders.

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

Issue 9: General Policy Issues

16. Comments and suggestions identifying any other issues related to the interface between copyright and AI are welcome. Specifically,

(i) Are there seen or unforeseen consequences of copyright on bias in AI applications? Or is there a hierarchy of social policies that needs to be envisaged that would promote the preservation of the copyright system and the dignity of human creation over the encouragement of innovation in AI, or vice versa?

> **EWC comment:**
> With indulgent astonishment we look at the attempt to once again play off copyright and innovation as well as copyright/authors' rights and data flow and data access against each other. Yet these qualities do not face each other as enemies.
> On the contrary:
> Since its creation, copyright and authors' right law has made innovation possible at all by explicitly protecting the work and investment of its authors of original works, whether of an artistic nature or intellectual achievements in science. This protection is intended to protect the investments that an author and his partners make in the development of those innovations that we call texts, books, knowledge.
> We therefore strongly advise against making copyright and innovation the opponents.
>
> We also advise against using copyright in a mindset as an obstacle to the free flow of data. It is not copyright usage agreements which inhibit data flows and the general public's access to knowledge and art! It's much more banal: administration effort and remuneration are the points where responsible parties would like to save themselves the effort and budget This may even be understandable at times, but it does not relieve us of the task of finding solutions that do not put only one side, the sources of knowledge and art, the authors at a disadvantage. In the long term, this disadvantage leads to an unfree, dependent and limited diversity of information and literature.
> Another point seems worth mentioning. Already now, digitisations for archives, but also developments of different AI, show prejudices towards certain groups of people, topics and cultural heritage. Digitisation and AI development partly increase discrimination against women and ethnic minorities. In countries whose consciousness is still shaped by the culture of colonization, governments curate memory through AI and digitalization. Accordingly, text machines «learn» context only from restrictively approved material. They are neither reliable witnesses of history nor reliable witnesses of the present.
>
> Instead of building an either-or in the solution finding, it would be the chance to think in «and». Copyright AND innovation - because authors and every author is the core of all innovation.
>
> Social policy therefore has one task: to focus on the AND.
> Not the « or ».

WIPO/IP/AI/2/GE/20/1 – consultation response by European Writers' Council – EU Transparency Register ID number: 56788289570-24

9

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

Issue 10: Further Rights in Relation to Data

(i) Should IP policy consider the creation of new rights in relation to data or are current IP rights, unfair competition laws and similar protection regimes, contractual arrangements and technological measures sufficient to protect data?

> **EWC comment:**
> There is currently no right of ownership for the (own) data, only a right of disposal. This is usually ignored, neither queried nor reimbursed.
> Here the protection of personal rights is required and the formulation of a sovereignty and ownership of the data that one produces oneself and often receives in return for supposedly free applications on the web. The term «intellectual property» or copyright does not apply here, as the data is docked to different areas of personal action and articulation - movements, purchases, preferences for people, media, cultural works, etc.
> This is the right to informational autonomy, which should be linked to personal rights.

(ii) If new IP rights were to be considered for data, what types of data would be the subject of protection?
(iii) If new IP rights were to be considered for data, what would be the policy reasons for considering the creation of any such rights?
(iv) If new IP rights were to be considered for data, what IP rights would be appropriate, exclusive rights or rights of remuneration or both?

> **EWC comment:**
> Data tracking comprises a number of different personal data and individual traces attached to the personality, which add up to a digital fingerprint. Sleep, exercise, change of location. Taste in music and film, taste in literature, political orientation. health data and also the circle of friends. Working hours and ethnicity, travel plans and appointments. Creditworthiness and even ability to concentrate. Family demographic details, favourite products, and how strongly one react to advertising, political news or fake news.
> These data are evaluated and sold. In some cases, the data is used against the person, e.g. in questions of creditworthiness or when calculating health insurance premiums.
> They therefore already correspond to a type of currency. The «conversion» is currently being done at the expense of the involuntary data donors.
>
> How this currency is to be legally assessed is, in our opinion, not by copyright law.
> Here, for example, a right to informational self-autonomy must be developed – which can also be a service or a currency. This service can be paid for by platforms and other users, for example, in return for payment or consideration. It must also be voluntary, per opt-in and self-autonom. A monetary system is also possible.

(v) Would any new rights be based on the inherent qualities of data (such as its commercial value) or on protection against certain forms of competition or activity in relation to certain classes of data that are deemed to be inappropriate or unfair, or on both?

Rue d'Arlon 75-77, 1040 Brussels - Belgium+32
(0)2 286 83 12
EWC-Secretariat@inter.nl.net
www.europeanwriterscouncil.eu

(vi) How would any such rights affect the free flow of data that may be necessary for the improvement of AI, science, technology or business applications of AI?

> **EWC Comment:**
> Again, we strongly advise against this suggestive thesis, which already contains prejudices, to put copyright in a mindset with the free flow of data as an obstacle.
> In the collection of personal data, contemporary humanity has the choice to decide according to Mill or Kant.
> A Mill's decision compares the weighing of consequences and the goal with the use of the means: Can we deny humanity the right to collect health data in order to prepare an AI for diagnosis, early detection or healing? On one hand, no. But what if this data gets into the hands of restrictive regimes, as the present situation suggests - and with what intentions, for example, do extreme-right governments use this sensitive health data?
> A Kant's decision is both easier and more difficult.
> It is not appropriate to interfere with the freedom and informational self-autonomy of a person and to value the individual less than the whole society.
> Here we are only at the beginning of an ethical debate.
>
> Fact is, that in the sense of constructive discourse, we should refrain from playing off copyright and free data flow against each other.

(vii) How would any new IP rights affect or interact with other policy frameworks in relation to data, such as privacy or security?
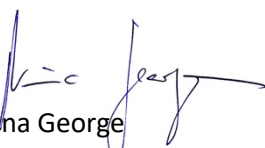
> **EWC comment:**
> In the development of applications, the focus is on questions of procurement - under legal and monetary aspects as well as those concerning the protection of personality - and the preparation of data, data protection against manipulation and misuse, and the provision of the necessary interfaces and responsible programming personnel. This includes necessary industry standards, but also ethical ground rules as well as secured budgets for the protection of the so-called «data», whether they are works or digital traces.
> This also affects liability issues, such as strict liability, compulsory insurance and who takes over these; whether contracts concluded with chatbots are valid, whether there is a liable violation of privacy; or damages that can arise from unilateral information monopolies, when an algorithm decides what is valuable and unworthy information.
> There is a lot coming up - it remains exciting.

We will continue to be at your disposal as a discussion partner.
With best regards

Nina George
EWC President