

Copyright and AI Reference Group (CAIRG)

Questionnaire: Copyright material as AI inputs – mapping exercise

Purpose:

The primary purpose of this questionnaire is to collect examples from CAIRG participants of how copyright materials are **currently** being used in Australia as inputs for AI systems, or could potentially be used as such **in the future**.

The Attorney-General's Department will use the examples provided for an **exploratory mapping exercise** covering:

- the various types of copyright material that may be of value as inputs for AI systems in Australia
- different ways in which developers and deployers of AI systems in Australia may seek to use copyright material as part of the AI value chain
- particular purposes to which these AI systems may be put, and
- practices and arrangements that may facilitate that use.

We have previously heard about the value of appropriate datasets for the training and operation of AI models and tools, as well as the connection between the quality of AI inputs and the accuracy, reliability and robustness of AI outputs. We know that AI models and the purposes for which they are designed vary significantly, and that inputs may be used at different stages of (and by different entities in) the AI life cycle. Through this questionnaire, the department hopes to gain a better practical understanding of the full range of current and potential future uses of copyright material as AI inputs in Australia.

To provide a broader context for this exercise and future engagements with the CAIRG, the questionnaire also asks participants to contribute their views on **what objectives Australia should be aiming to achieve** where our copyright framework intersects with the development, deployment and use of AI technology.

The evidence base that the department will compile from responses (in deidentified and consolidated form) will support future engagements with the CAIRG to identify key problems and uncertainties at the intersection of AI and copyright in Australia, explore potential solutions (legislative or non-legislative) to address these issues, and articulate key principles and objectives the Government could be guided by in considering available options.

What we are seeking from you in this exercise:

Questions 1 to 3 ask you to confirm whether any information you're providing is commercial in-confidence or shouldn't be shared within Government, and to provide some information on the capacity in which you're responding to the survey.

Question 4 requests your view on the objectives Australia should be aiming to achieve at the intersection of AI and copyright. We would like you to give us your high-level vision of what success would look like in a 'national interest' sense. While other parts of the questionnaire are focused solely on AI inputs issues, your answer to this broad, overarching question may also address AI outputs issues.

For the primary mapping exercise, addressed through Questions 5 and 6, we are seeking the following:

1. Examples of how and why AI developers and deployers in Australia might use copyright material as inputs in the development and operation of AI systems.
 - We are interested in all types of AI systems (e.g. 'narrow' AI models as well as foundation/general purpose models) that may use copyright-protected material for any purpose, as well as uses at different stages of the life cycle for AI applications (e.g. pre-training, fine-tuning of pre-trained models, Retrieval-Augmented Generation (RAG)).
 - Examples could relate to uses that you know or believe to be already occurring in Australia; uses that are already happening overseas that could potentially also occur in Australia; or theoretical or proposed future uses that you believe could be part of the future of AI development and deployment in Australia.
2. Examples of types of materials that may be copyright-protected, or collections of materials, (copyright material) that are of known or potential value as inputs for AI systems in Australia.
 - Examples of copyright materials used in AI development in Australia can include Australian-owned copyright material and/or overseas-owned copyright material.
 - Any relevant information you are able to provide on the scale/volume of materials being used or could be used, and on who may own and/or may control the copyright in such material (in general industry terms, rather than individual owners' identities). This information will help provide context for the department's mapping exercise by illustrating the opportunities and potential challenges surrounding use of materials for AI purposes.
 - If you would like to provide any additional information on non-copyright interests in materials that are used (or that would be helpful to use in the future) in AI that you think are relevant to the CAIRG (e.g. moral rights, performers' rights, Indigenous Cultural and Intellectual Property rights in copyright material), you can do so in sub columns Q5(F) and Q6(E).
3. In relation to the above, information about the practices and arrangements through which these uses of copyright material may be facilitated (now or in the future), as well as about any issues affecting these uses.

- Practices and arrangements could include (for example) commercial arrangements such as direct licensing agreements, use of materials published under open licenses, exceptions under copyright law, or the use of proprietary training material.
- We would welcome information on practices and arrangements that you know or believe are currently being used in Australia, or that you believe could be developed or utilised in the future.
- Issues affecting uses of copyright materials as AI inputs could relate to industry practices, or legal or regulatory uncertainty, among other things.
- We also welcome your views about the impact or utility of any voluntary non-legal mechanisms to control if or how copyright works can be used in the training and development of AI systems (e.g. opt-out regimes, 'Do Not Train' registers, or other technical mechanisms such as .txt exclusion files).

You may provide as many or as few examples as you like.

The more specific you can be in the examples you give, the better, as this will help us map out the Australian AI-copyright landscape in its full diversity. However, we do not necessarily expect examples to identify particular organisations or other entities, or to reveal sensitive or confidential information. Examples could be drawn from your direct knowledge of industry practice, from press releases or media reports, or from your expectations or aspirations about how the use of copyright materials as AI inputs could develop in the future.

Respondents will approach this exercise from different perspectives and with different areas of expertise. If you are not able to provide information about one or more aspects of an example you are providing, feel free to leave parts of the questionnaire blank.

What we are not seeking from you in this exercise:

Information or views on the following issues, which could be addressed in future engagements with the CAIRG.

1. Uses of copyright material (Australian or otherwise) as AI inputs outside of Australia, and therefore not subject to Australian copyright law.
 - We acknowledge that there may be some grey areas here – for example, if one stage of training for an AI system occurs in a foreign jurisdiction and another stage occurs in Australia. If you provide an example relevant to the development and deployment of AI in Australia in which some uses of copyright material as AI inputs have occurred outside of Australia, we would appreciate it if you could make it clear, to the extent possible, which uses are occurring in Australia (and which are not).
2. Legislative, regulatory and policy approaches to the use of copyright material as AI inputs in international jurisdictions.
3. The commercial value of particular types of copyright material as AI inputs (in Australia or otherwise).
4. Specific suggestions on how Australia's copyright law could or should change (e.g. to facilitate AI development in Australia, provide greater protection or transparency for rightsholders, or resolve other perceived gaps relating to AI inputs).
5. With the exception of Q4, issues related to AI outputs.

Questions:

Q1 – Does your response to the questionnaire contain any commercial-in-confidence information that you would like to be excluded from material the department prepares for circulation to CAIRG participants and/or publication, even on a deidentified basis? (As noted above, we do not necessarily expect examples given in responses to identify particular organisations, including your own, or to reveal sensitive or confidential information.)

Yes – entire response Yes – particular commercial-in-confidence information **[square bracketed and highlighted]** in questions Q4 through Q7 No

Q2 – Do you consent to information provided in this questionnaire response being shared on a confidential basis *in original, identified and unconsolidated form* with the Attorney-General, other Australian Government Ministers, and/or relevant Australian Government agencies for the purposes of AI policy development?

Yes No

Yes, with limitations – please detail:
.....
.....

Q3 – Please indicate the capacity in which you are responding to this questionnaire (more than one may apply):

- AI Developer (i.e. an entity that designs, codes, or produces an AI system, for its own use or for use by a third party, as a baseline model)
- AI Deployer (i.e. an organisation that operates an AI system itself, or for use by third parties, that it may have fine-tuned to meet particular needs)
- AI User (i.e. a person or organisation that uses a third-party deployer’s AI system)
- Creator of copyright material – please indicate which type(s) of material:
- Rights-holding individual/organisation
- Industry body representing one or more of the above – please indicate which: Australian Institute of Architects. At present we are aware members that members are AI users and also creators
.....
- Copyright collecting society
- Other – please specify:
.....
.....

Q4 – What high-level objectives do you think Australia should be aiming to achieve where our copyright framework intersects with the development, deployment and use of AI technology? *Please limit your response to 100-300 words.*

Your response could cover (among other things):

- Goals or aspirations for one or more industry sectors, including how the goals and aspirations of various sectors could be balanced
- Broader economic, social and cultural considerations, and/or
- How these goals, aspirations and considerations relate to the existing policy objectives of copyright law (as you understand them)

The high level objectives should be as follows:

1. **Establish a demarcation in copyright between creative work being used to inspire further creativity versus technology being used to systematically reproduce or synthesise creative work and ideas by “training” artificial intelligence software.** When creative work is used to inspire further creativity, the second creator uses a range of subjective processes to create their work these involve:
 - philosophies
 - values

- affective (emotional responses)
- life experiences (direct or vicariously through association with an proximity to others)

In contrast, artificial intelligence samples the outputs of human beings’ creative processes. Artificial intelligence does not create through subjective processes and samples the outputs of creative processes. What artificial intelligence powerfully does is systematically and efficiently sample the work of others, performed rapidly to algorithms.

This demarcation is important as people who engage in creative processes will often say that the work of another had inspired them – examples include writing, painting and sculpture, architecture, and music. Motifs may even be borrowed and used to the extent of creating a legal dispute. A famous example is the 1975 pop song. “All By Myself” by American singer-songwriter Eric Carmen, of which the verse is “*based on the second movement (Adagio sostenuto) of Sergei Rachmaninoff’s 1900–1901 Piano Concerto No. 2 in C minor*”¹. clear ground rules need to be established that extend beyond traditional notions of copying to use material to train Artificial Intelligence. CDlear ground rules need to be established that extend beyond traditional notions of copying to using material to train Artificial Intelligence.

2. **Establish an attribution requirement.** Much like the issues associated with attribution in academic writing and the ethical offence of plagiarism, requirements should be established about when work should be attributed to the input sources. If AI, for example, is used to sample a large body of work, or a large percentage of the detail of one work (this would vary according to the type of work) then the question is raised as to what attribution requirements should be established.
3. **Establish a plain language disclosure requirement for software and online storage providers about how created material will be used.** We have heard from members of the architecture profession that some software tools they use extensively in practice rely upon the cloud service of the software provider to store their files without an option to use locally hosted storage (i.e hosted by the software user). They report not being clear about licence agreements and are concerned that the intellectual property rights in work created using these software packages including files that that are held in the software vendors’ hosted cloud storage do not exclusively vest in them.
4. **Establish consenting and opting-in requirements for creative material to be used for training AI systems.** Creators should retain agency over their created work. A creator may choose to teach master classes in their work and style to students for a fee. Their material should not be available without their consent and knowledge to train artificial intelligence systems. This may particularly apply to software as in point 3 above.
5. **Establish a requirement for AI systems which scrape publicly accessible material to recognise copyrighted and attributable work.**

Q5 – In the following table, please provide examples of how (to your knowledge or belief) copyright material **has already been used, or is currently being used**, in Australia as inputs in the development, deployment and/or operation of AI systems. You may provide as many or as few examples as you wish. *Note – the **Example** the department has provided is theoretical only, and is not based on any known real-world activity.*

A) Describe the AI system (e.g. purpose and uses; intended user-base; commercial/non-commercial application). <i>Examples do not need to identify specific individual developers/ deployers/users.</i>	B) What kind(s) of materials, that may be copyright, were used as inputs? <i>Please be as specific as possible, and list as many kinds of material as are relevant. Examples do not need to identify individual creators or rights holders. If known, please provide any quantitative detail on the scale or volume of materials used.</i>	C) How in particular were these materials used as inputs for the AI system? <i>Responses could cover the stage(s) in the AI system’s life cycle in which the material was used (e.g. pre-training, fine-tuning, Retrieval-Augmented Generation) and the way in which this use contributed to the AI system’s functioning.</i>	D) What practices and arrangements have supported developers and/or deployers’ use of these materials as inputs? <i>Responses could address any industry practices and/or legal arrangements under which developers/deployers have accessed and used material.</i>	E) Are there any issues (e.g. with industry practices, or legal or regulatory frameworks) that are currently affecting these or similar kinds of uses of copyright material as AI inputs in Australia?	F) Additional information/ comments. <i>e.g. links to external sources or research, relevant information not covered in previous columns</i>

¹ Source: https://en.wikipedia.org/wiki/All_by_Myself#:~:text=All%20by%20Myself%20%2D%20Wikipedia

<p>Architects report that they use readily available online AI tools that create designs and images to create renders and concept designs of buildings and structures.</p>	<p>Images of buildings where both the building design was subject to copyright and the images created by professional/ architectural photographers as well as 3D renders of buildings not yet built which may have also been subject to copyright.</p>	<p>The software being used relies upon word prompts to create concept designs or renders of a new building. This can also enable the work of one single architect to give specification to the AI software to create a render or concept that could be based upon multiple samples of that architect's work (the architectural style). However it may be possible to prompt the AI to focus on one specific design or building - which increases the risk of a copyright infringement.</p>	<p>Images of buildings appear throughout the internet.</p> <p>Images also appear on the internet which include information about the provenance of the architectural work and also the image creator (photographer).</p> <p>In the instance of our own organisation we have publicly accessible images of award winning architectural work (from our annual regional, states and territories and national awards program). Members consent to those images being available for specific purposes including our other communications, marketing and social media outputs.</p>	<p>It is important for architects to be clear about protecting their own copyright material and intellectual property as well as being aware of when they might be infringing the copyright of others.</p> <p>There needs to be greater clarity about how AI is trained and used and the copyright risks attendant upon training and using AI.</p>	
--	--	--	--	--	--

Q6 – In the following table, please provide examples of how (to your knowledge or belief) copyright material **could be used in the future** in Australia as inputs in the development or deployment of AI systems. You may provide as many or as few examples as you wish. *Note – the Example the department has provided is theoretical only, and is not based on any known real-world activity.*

<p>A) Describe a potential future AI system (e.g. purpose and uses; intended user-base; commercial/non-commercial application).</p> <p><i>Examples do not need to identify specific individual developers/ deployers/users.</i></p>	<p>B) What kind(s) of materials, that may be copyright protected, could be used as inputs?</p> <p><i>Please be as specific as possible, and list as many kinds of material as are relevant. Examples do not need to identify individual creators or rights holders.</i></p> <p><i>If known, please provide any quantitative detail on the scale or volume of materials that could be needed and/or available.</i></p>	<p>C) How in particular could these materials be used as inputs for the potential AI system?</p> <p><i>Responses could cover the stage(s) in an AI system’s life cycle in which the material could be used (e.g. pre-training, fine-tuning, Retrieval-Augmented Generation) and the way in which this use would contribute to the AI system’s functioning.</i></p>	<p>D) Are there any issues (e.g. with industry practices, or legal or regulatory frameworks) that could affect these or similar kinds of uses of material, that may be copyright protected, as AI inputs in Australia?</p>	<p>E) Additional information/ comments.</p> <p><i>e.g. links to external sources or research, relevant information not covered in previous columns</i></p>
<p>Architects use software tools throughout their daily practice to produce documentation of buildings down to construction details at a 1:1 or 2:1 scale.</p> <p>Construction details may also involve innovative approaches to creating performance solutions.</p> <p>It is already common in daily architectural practice with public and community buildings (education, health and community infrastructure), commercial and retail buildings as well as multistorey apartments for designs and all documentation to be uploaded and integrated into a digital three dimensional building information model. These software packages and the files generated may be hosted in the cloud.</p> <p>It is conceivable that software vendors and cloud storage providers may seek to use the created work to train AI building design systems. This may risk infringing on original architects’ work.</p>	<ol style="list-style-type: none"> 1. Entire building designs including novel methods for creating end use solutions and novel methods for constructing a building or component elements. 2. Jurisdiction specific specifications for planning (e.g apartment design guidelines) other planning provisions and then the applicable provisions of the National Construction Code. 3. Australian Standards that pertain to the design of buildings and outdoor spaces (e.g AS1428.1 Design for Access and mobility or AS 3959 Construction of Buildings in Bushfire-Prone Areas) which are copyright to Standards Australia. 4. Specifications for building performance such including energy efficiency, engineering, structural and services (e.g. ventilation) computations. 	<p>Copyright building designs could be used to train the AI software together with planning, national construction code, Australian Standards and other performance computations. Additionally, algorithms from systems to evaluate building could be incorporated into these systems.</p> <p>The end result would be software which could perform a lot of work undertaken by architects and other building designers and which would be of considerable commercial value.</p>	<p>Industry practices should require open disclosure of how specific types of content have been produced such as concept designs. When a consumer or client approaches an architect or any building designer for a design – there should be transparency about the provenance of the design.</p> <p>This needs to be balanced off against the benefits and assistance provided by AI based or augmented software. Innovative use of technology will continue despite the resistance or even perceptions of redundancy by any one creative group.</p> <p>What is important is that original work created should have its copyright respected and clear ground rules need to be established that extend beyond traditional notions of copying to use material to train Artificial Intelligence.</p>	

Q7 – Are there any other comments that you would like to provide on:

- a) things the department should consider when using questionnaire responses to build a broad evidence base on current and potential future uses of copyright material as AI inputs in Australia, and/or
- b) how the department could approach future requests for information from the CAIRG to build on this initial mapping exercise in relation to AI inputs?

We importantly note that this is an exploratory survey and the information prepared here can be considered neither exhaustive nor ‘rigorous’. It has been prepared from some ongoing discussions at meetings with small groups of members. However, there should be further in-depth descriptive and analytic research conducted from much larger survey samples undertaken directly with architecture practices as well as with software vendors. Such research should also include qualitative methods of in-depth interviews, focus groups and even field work such as process observation in order to form a thorough understanding of how creation in fields such as architecture and building design is occurring and where or how copyright might be infringed.