1. AI and word/image processing systems are complex, powerful tools which I view are an extreme extension of programs that compact and automate processes not unlike the functions and brushes of an image editing software. Unlike such softwares, which have a very specific purpose, AI programs are fully capable of being abused to not supplement works, but to create them wholesale; and furthermore, the functions of an image editing software work mainly in the context of the image in its singular composition, while an AI must first consume other materials and derive its own context, and those materials currently are not limited to a specific pool of media from which it is allowed to take. A human is capable of borrowing context, technique, and inspiration from various media at will, but is capable of manifesting new ideas from said work, which a computer currently is not. By extension, AI trained on particular works and authors of works who have not submitted a willingness to be sampled may find that their work now resembles an AI's creation, rather than the other way around.

2. In the realm of the book industry, AI material - while currently inoffensive in its structure, as most programs are trained to be mild - is technically indistinguishable from written work created by a human, unlike image or audio generations in there are still some identifying flaws or styles. As time goes on and word processing LLMs become capable of more complex, consistent generation - potentially understanding narrative or genre - I fear that some authors may choose to pay for, acquire, or even train themselves a processor which is capable or producing a written work at a fraction of the time it would take to produce themselves, then submit them to agents and publishers who are none the wiser to the text's origins.

4. 5. There is no way to ensure international consistency of any technological mandate or copyright. I believe that, optimally, most nations would adopt legislation which limit the input of AI training pools to works which have been sold or been created for the exact purpose, closely observe companies which sale or allow the public use of AI processors, and demand that AI generated content be clearly marked, even if personally generated by the author's training, regardless of whether or not an attempt to procure the rights to that content was made. Similar works have been created in the past, present, and for as long as humans will create, but there is no novelty in the processing in an AI, and thus there is no right to ownership.

6. Text on a webpage, or in a digital document, is seamless to scrape off the public internet and put into a processor. Less complex AI can be used to collect content in this way, curated by humans or otherwise fed directly into a larger LLM, or this content can be sold by data-collecting companies in regards to more specific goals. (advertisers seeking consumers by demographic, ect.) More abstract mediums can be sorted by humans, or scraped by apparent popularity in an online forum or individual interest. There is nothing stopping the developers of LLMs from taking any works, copyrighted or otherwise, and feeding them into their machine, as the expected and curated result will likely have not borrowed enough any one part of its training to be indistinct. I am unsure of the extent that developers commission or create the material for purpose of training their AI.

7. I understand that LLMs create combinations or phrases, words, or compositions through analyzing batches of curated content and, through a process of probability and association, can reproduce apparently novel prose by taking a topic and then accessing its memory for likely (more common) sentences and words to fill in the gaps. This is largely dependant on the training it recieves, taking a sum total of any context it can derive from something and producing a neutral, inoffensive, ultimately muddied piece of content. I assume it works similarly for other mediums, taking the lump sum of what is put into it and attempting to reproduce an approximation of the content you ask for it. An AI has no concept of its own 'correctness', and cannot identify what to us would be obvious issues in its output, and must be curated so as to have parameters that make sense humanly. If an AI forms a 'habit' or otherwise is seeded an undesirable gene, a developer may not be able to identify the exact source of the error with the sheer volume of material used to train it, or the lack of understanding in the exact mathematics as to how the AI came to such a conclusion, but may still be able to simply discard the resulting errors. If the machine is working as it is designed to, the errors are only just that, and they will be processed out or ignored. The problem with having a large dataset for training, while simulating novelty itself, is the AI may still be able to reproduce individual parts of its training when asked, and may be curated so that individual sets of training material were favored over the rest, resulting in a recognizable common output.

8. I do not believe any unauthorized use of copyrighted works should

constitute fair use when used to train an AI. A learning algorithm is not a person, but a tool, and the product which exits it, if desired to be considered fair itself, must also have incorporated parts which the developer themselves have the right to use. A movie cannot use a sample of a copyrighted song without first obtaining the rights from the publisher, even if the song itself was altered in some way, or appears for only a short duration over the length of the film. Just the same, the product of an AI is only whole and belonging to its user if said product was developed from a training model that infringes on no individual's or company's rights. The output of an AI is less important than the parts that were used to produce it. A dataset containing unauthorized content may not be used commercially, and if the content is authorized, that product be marked as having been generated by a learning system.

9. Authors must Opt Into having their work be materials for training models, and furthermore, the companies or idividuals which collect this data, to sell or to use, must make this information clear and official. There is no reasonable world in which Opting Out is a viable option, as it to the profit of certain companies to obfuscate the process to do so if such a thing were to exist, working to make it as difficult or confusing as possible so as to reduce the loss of potential material. It should not fall to the author to navigate that process, but to the party which is seeking out data to train with. Violations of said process by the discovery that a model contains unauthorized works should give the author the right to demand the product in which their works were involved be remade without their involvement, or settle the infringement with the owner of said AI through rightful sale of work or by an official share of ownership of said product resulting from the generated content. The materials used for training should come from sources which themselves have ownership of said media and themselves authorize the use of their intellectual propery for training purposes.

10. Licenses to the copyrighted content for use of training an AI model would likely be recognized through intermediary systems or companies which purchase intellecual properties, or the rights to intellectual properties, for sale to the respective parties. Original creators the works should be clearly informed that their content may be sold if their publisher or copyright holder would have intentions to do so, even to an intermediary organization. Royalties may be offered depending on the agreement reached between the copyright holder and the intermediary, or the content may be sold wholesale. Once a work is sold, there would not be an ability to opt out of the agreement apart from a case-by-case basis or an infringement of the original owner's fundamental rights.

11. There is no guarentee that the organizations which mediate the obtaining and sale of training-legal datasets would follow procedures which ensure that the intellectual rights of the original owners are honored. The curators of the data which will be used to train an AI model are responsible for the integrity of their data in regards to their legal right to use that data, and the contents of these datasets must be properly and wholly logged and submitted for authenticity. These can, of course, intentionally ommit data which might infringe upon copyright laws, and as it stands I do not know how one would be able to assess the results of a trained AI for fraudulent content without having access to the data which it learned from in the first place.