1: It's not really relevant? This is like asking how industrial looms making textiles efficiently (and sometimes in designs that humans didn't intentionally input) might impact creators, copyright owners, etc etc. It doesn't matter, because the nature of the device isn't the sort of thing that should have legal ramifications, outside of, say, job retraining and a robust social safety net for those textile workers who have been displaced by automation taking their jobs.

2: Again, copyright isn't really a concern on this issue. The model itself doesn't contain any infringing data. The model could be overfitted, and so tends towards outputs that are infringing, but it's not obvious that this is a problem with the model, legally speaking. It could be the *prompt* that produced the overfitted output that's considered infringing. And even then, this effect is relatively small and can likely be avoided by proper care of training material. (Which probably *should* be regulated to some extent, but not to the degree that many people wish. If we make data scraping contingent on people giving their consent, things like google will fall apart.)

3: https://arxiv.org/abs/2301.13188

4: Not to my knowledge, but you should completely ignore other countries in this area. They often have very poor standards for internet regulation and stifle far too much.

5: Yes, in the sense that copyright in general is broken and has been broken for years. And part of the issue we're having is that we're trying to use this broken system in areas it just fundamentally wasn't meant for (the internet). It needs to be fundamentally rethought - many people using copyrightable elements aren't corporations with lawyers, fair use needs to be expanded, and if a corporation fails to keep a work on offer for those who wish to buy it, the copyright should lapse. But none of this is going to happen. And given this the best option is to regulate generative AIs specifically (or even specifically GANs) and to leave copyright alone entirely, as touching it without massive reform will make everything worse.

6-6.4: Let's just be honest, it's usually piracy. At least for academic researchers. They don't have the resources to partner with people for large quantities of training data and public domain works are not anywhere near sufficient. Like, sci-hub is as popular as it is as a reason. Academics pirate when they need to. And for this they need to. They'll probably keep the data on hand in a file somewhere just so they can reference it, but they're not *using* it, outside of the training.

7-7.4 I'm not an AI researcher, but from my understanding of the subject as someone slightly adjacent to the field, it's near impossible to "unlearn" things without unravelling the entire model, and you can't tell if a model has learned from a specific piece of data unless overfitting is present. As for how inferences are stored, this is sort of a bad question. Suppose I gave a very advanced neural net the following data set: {(1,1), (2,2), (3,6.1), (4,24), (5,120)}. It might return a program that represented the factorial function in C++. (As neural nets are "models that make models") No inferences are "stored" anywhere in this model. It went from a collection of discrete points to a broader "worldview" that largely aligns with the data we fed into it.

8-8.5: It's fair use in every sense. Google Books is very clear on this. There's just nothing here. Neural nets make models. Using training data is *making a model*, it's not creating a work of art. The end result, the model, then can make art given a prompt. But that's different. Collecting the data for training is the exact same analysis, web scraping is legal. And the commercial nature isn't really relevant here, Google

doesn't make direct money off of Google Books Search, but makes quite a lot indirectly. And, again, training a model is different from the model then creating art. You're collapsing two steps to one step. Training the model is *inherently* fair use. The model then exists and *might* churn out infringing content depending on the prompts. But that's a separate issue.

9-9.5: Copyright owners should have limited say on whether their material is used for training AI work, just as they have limited say whether other data is being used for web scraping. But what say they have must be opt out. Implement policies similar to what is used for web crawlers generally.

10-10.5: This shouldn't happen, it would break the internet if consistently applied. Google could no longer index websites.

11: Well, one issue is that you can't get in touch with many of the copyright holders, which is the issue that happened with Google Books. It's almost like this is a startlingly bad idea.

12: Not unless overfitting happens. That's just not how neural nets function? We don't understand them well enough to tell you these sorts of things about them.

13: Pretty bad, which is why nobody has asked Google to do this.

14: Again, I just want to reiterate the two big points. Data scraping is foundational to the modern internet. You cannot require it to be licensed or opt in. That would kill everything. Secondly, training a neural net is *radically* transformative, as it gives you, in effect, a mathematical model. The then existent neural net will then go on to do similar things as the original artist, but that's a different issue. You need to separate those two stages.

15-15.4: I mean, theoretically, if there was no way that the people could opt out. But the issue is that the narrative around AI at the moment is so doom and gloom, for no reason, that hearing this will cause people to freak out and act irrationally.

16: Again, it would be nice to let them know, just to be informed, so long as they couldn't do anything about it.

17: I have no knowledge of this.

18: Yes, but it takes a substantial level of human intervention. Mere prompt engineering is not sufficient. Prompt engineering + editing in photoshop? Probably.

19: Probably not.

20-20.1: Status quo is probably fine.

21: No, but neither does it permit much of the current things that are copyrighted today.

22: Yes. I'm not certain how similar they'd have to be. Is just having Spiderman in a picture enough? Maybe. But the issue here is the output, not the neural net.

23: No clue, would have to be ruled on by courts.

24: Oh, no, bad question. The training data is irrelevant here. It's the output that matters. The neural net doesn't have any infringing data, by definition. It can't. It isn't the sort of thing that can have infringing data. That would be like saying that you infringe on the copyright of Spiderman by having the concept of "Spiderman", rather than your *drawing* of Spiderman being the thing that infringes. It's definitely the latter. Whether or not this is the issue with the prompt or the neural net is tricky. But it's not the training data.

25: Really unclear? Probably the system developers or the end users. You probably need legislation to clear this up. If you sanitize your training data to avoid overfitting you can probably pin it on the end users. But you still need laws in place to make this rigorous.

26: It does not? Completely unrelated to training a model.

27: I think you hit the big points here.

28-28.3: This is a good idea, yeah. I think ideally this should be pinned to the user, but this would be hard to enforce. Probably add something at the system dev level. And the idea situation is a fine, as well as a fine for removing it.

29: Not my area, but these tools will always be fundamentally incomplete as they're lagging behind the best work.

30: Not sure? I think if they're a public figure it's probably find, but if not you might want to pass a law outlawing it.

31: There should be a floor law regarding AI privacy around, say, your voice and likeness. I think those things are pretty common sense. Much of the rest is really more nuanced.

32: No. You cannot copyright style.

33: Honestly not sure.