I am commenting as a private person, a computer science student, a musician, the spouse of an artist and a general humanist.

I will answer the general question number one and then discuss the training of AI models.

On the general question

My view is that the risks are much greater than the benefits. Most of the benefits are economic in nature, such as allowing businesses generate cheaper marketing campaigns, cheaper illustrations, cheaper web content, cheaper musical or filmographic products, etc. From the human perspective, however, cheapest and fastest production of content such as text, images, sound and video is not of benefit if the quality and meaningfulness of it declines, or if the leap in automatization of creating said content creates unemployment and social inequality. Furthermore, I do not think automatization of creative work is a goal at all from the human perspective, since I believe it is generally agreed that creative work is the most enjoyable and mentally rewarding kind of work, which automatization should specifically allow people to do more.

Additionally, generative AI technologies allow for great societal disruption, by making the creation of misinformation much easier and more efficient. Even though it has always been the case that one should not trust everything on the internet, before the latest AI technologies people still could trust the authenticity of most photos and videos and could trust that generally any text on the internet is written by a human, even if not a benevolent and well-informed human.

On training

First, I need to address the terms of â€œtrainingâ€ and â€œlearningâ€ themselves. Some people use rhetoric where they equate training an AI model to the learning process of a human. However, I believe that this is completely false and misleading, and should not be the basis of any ethical or legal reasoning. Training an AI system is not like the learning process of a person due to the following reasons: The structure of a generative AI model, called a neural network, does not resemble the structure of a human brain, if not for the fact that the components of both are called â€œneuronsâ€. The learning process of a human and the training of an AI model are not alike: a human does not learn by being fed millions and millions of images or text examples which themselves automatically translate into changes in the neurons, a human learns by slowly and deeply analyzing and reasoning about a much smaller number of examples, making decisions about the value of the observations and memorizing selected and meaningful pieces of information.

Because we live in a capitalistic world, the market impact of the learning must be also considered. A human training themselves is slow, unlike the training of an AI model. A human that has learnt to paint, for example, can paint only so many paintings in a period of time, unlike an AI model. A human that has learnt to paint can only paint themselves, unlike an AI model that, as a piece of software, can be copied unlimited times and therefore cause very great changes to the market value of paintings. Because of this fact, a human and an AI model cannot be compared from an economic perspective either.

Then I will discuss the problematics regarding the training material of AI models. I am aware that current copyright law is mostly focused on copies, as the name suggests, preventing people from making direct imitations of works. I am not going to argument so much based on the current copyright law, but based on what is just, and therefore what the law should be, since the ultimate purpose of the justice system is not to follow arbitrary laws, but to bring justice.

I strongly believe that any material should not be allowed to be used to train AI models unless explicitly permitted by author. In other words, I support the model that copyright holders should be allowed to opt-in to allow their works to be used to train AI, not have to opt-out to disallow. I would not like to limit that to copyrighted works, however. For example, I am under the impression that internet forum discussions are not copyright protected,

and they do not need to be, since there would be almost no way for anyone to profit by simply copying and distributing those forum discussions. But in AI training, those same internet discussions can be exploited for profit, and I do believe many writers would not have posted their writings on the internet if they would have known that those writings would be used for the profit of a third party. And the case is much stronger for copyright protected works such as books, paintings, and music recordings. For convenience's sake, I will call all of these creative workers such as artists, musicians, authors, actors etc. 'creators'.

The current situation of creators is deeply unjust. Firstly, the training material is an essential part of the AI model, arguably even more essential than the structure of the model. The current generative AIs simply would not be possible without the training material. The creators of the works used in training are to thank for making these things possible, but in reality, they are not getting any of the tremendous amounts of money involved in the generative AI business. Secondly, the AI models create direct competition for creators. By every work a creator publishes, and publishing is a crucial part of making a living in creative fields, they are directly benefiting their competition. Creators must choose between giving up their job or helping their biggest competitors to erase their whole field of work from existence. This is an ethical conflict so great that the law must somehow change to consider this injustice. Any other outcome would be a deathblow for all creative fields, leaving the world with very severe unemployment and stripping people from being able to enjoy novel non-AI-generated content. Additionally, legislation has to protect the creators whose work has already been scraped and used in training, optimally forcing the AI companies to abandon their unfairly trained models, possibly even with illegally scraped data. After that, they could train new models with licensed data. Even the interpretation of the past Google Books -trial has to be reviewed, since the technological advancements have completely changed the ways data masses can be analyzed, used and profited from.

In addition to works and non-copyright-protected content such as forum posts, people's identity and personalities should also be protected from being used in AI training. Again, this is not necessarily a conclusion so clear if considering only current law, but it is ethically very intuitive. How would it affect people mentally if physical identity could be copied with no limitations? Physical identity and recognizing other people's identities is such a fundamental part of the human mind, that allowing the free digital reproduction of people's identities would be a serious mental and societal disturbance. There should be strict regulations on training AI models on people's faces, gestures, voices etc. Using persons as training material should also require explicit opting in. I also consider the style of an artist, writer, musician etc. to be a part of their personality and identity, so it is in a way a violation of identity to machine-copy a person's creative expression. And even if the data is not used to mimic a person's appearance or voice, I don't think businesses should be able to profit from people just existing in this world of digitalization, where communication via the internet is necessary.

Of course, achieving all of these previously mentioned goals would require setting strict rules for transparency of the training data. Large technology companies are not well known for respecting their users' data or ethical standards, so there has to be some way to monitor what is used for the training data.

Artificial intelligence companies and lobbyers of course would answer all my arguments saying that strong regulation would 'hinder the development of AI', 'give the lead to China' or be impossible now after the damage has already been done. But I do not think humanity should make sacrifices so great just for the sake of AI companies' profits. Even though lobbyers have made it sound so, development of generative AI is not a glorious and important mission of humankind, it is only a mission of money-making by creating mimicry machines. Cancer will not be cured with the help of language models. Climate change will not be fought with the help of image-generators. China will not get a lead by 'the west' not wanting to allow AI companies to profit at the expense of creative fields and citizens identities. China would only get 'a lead' in destroying its artists' livelihood and citizens' trust to information. The argument that It would be 'too late' to regulate AI is of no value; the regulation of new technologies always has to be done afterwards. 'The west' can regulate these harmful ways of applying AI technology while reaping the benefits of applying it in more important, ethical and socially healthy ways.