OpenAI has asked the Trump administration to consider training on copyrighted materials as fair use (mentioned by Professor Levitan in the Hillel Place AI discussion on 3/27). Fair use is usually for material that is creatively transformative, which arguably is not done by the process of converting the copyrighted data into tokens because it's not using the concepts, ideas or face appearance of something to come up with something new based on a natural formulation of ideas. Rather, the data is being converted into something technical in order to have the AI model use the tokens to generate something similar to the data without doing so as a part of the creative process. However, artists and human creators are not paid or compensated for this use of their data or paid licensing fees to use their data in an authorized manner. As such, AI is considered to be a “soulless” imitation of human created art that doesn't compensate the originators of the tokenized data by human artists and creators. Therefore, policy recommendations will be guided by the following question: how can AI policies be shaped, reformed and created to credit and compensate artists for AI-generated work that was generated with training from their data especially given the direction that the US is heading in considering such work to be fair use and the previous support of the US in the advancement of AI and seemingly wanting to be a world leader in AI research and development? (Possibly consider legal AI policies from other countries that are more in favor of attributing artists’ work, paying licensing fees to artists to be given permission to use their work in this manner and/or compensating artists when their work is used to train models)

* One policy recommendation is to have other countries around the world, including the US, adopt a similar system to the EU’s AI Act for regulating AI and amend it to label AI programs that scrape art, writing, lyrics and/or music from human artists on the Internet as high risk. With this, there are a few recommendations:
  + Require AI programs to be transparent about the artists/originators on which the model is trained on/learned from in order to generate the created output
  + Make registration obligations such that AI programs cannot be used to generate certain creative content unless originators give permission for their work to be used to train AI models for such use and are compensated for licensing fees
  + Make publicly accessible documentation on how the AI model is trained to do these things
  + Make data quality requirements such that AI programs are not allowed to be trained on copyrighted material