

## Is Scraping Stealing?

- “Scraping” is the Process of Training Large Language Model (LLM)/Generative AI models by “Ingesting” large amounts of Data, usually from the Web, including Books, Designs, Pictures, Songs, Code, etc., where much of such Data is Copyrighted and not Licensed.
- Creative Professionals, and those that support them, including Authors, Songwriters, Photographers, Designers, and Coders, argue that this unlicensed copying, reproduction, and creation of unlicensed Derivative Works is a violation of Copyright – hence, Stealing.
- AI companies argue that such unlicensed ingestion, copying, reproduction, and creation of “new” works is a “Fair Use,” despite the commercial nature of the use, arguing such use of copyrighted works is “Transformative,” relaying on the Google search engine decision that found indexing for search engines transformative and therefore a Fair Use.
- Creatives differentiate Google and other Fair Use decisions and point to the output of AI as commercial and where all or substantially all of a photo, song, book, or code is generated by the AI generator, to refute the “Fair Use” argument, particularly after Warhol