

May 30, 2023

## **Innovation and Regulation: Finding the Balance for AI**

### **Publications**

In the fast-paced world of artificial intelligence (AI), countries around the globe are struggling to create suitable legal frameworks. Israel, renowned for its dynamic tech industry, has frequently found itself following the lead of other jurisdictions, particularly Europe and the United States, and adjusted its regulatory strategies to match global trends, rather than proactively adopting its own internal guidelines. This pattern seems to be persisting with AI legislation, where a clear position has yet to be declared by the Israeli legislator. However, in an active effort to create some certainty for developers of AI products and to boost innovation within its world-renowned tech sector, the Israeli Ministry of Justice published a few months ago an opinion on the complex relationship between copyright law and machine learning (ML).

Natural Language Processing (NLP) is a branch of artificial intelligence that focuses on the interaction between computers and humans through language. The learning process of NLP involves training algorithms on large volumes of text data, enabling them to understand, interpret, and generate human language in a meaningful and useful way. The necessity for vast databases in this process stems from the complexity and variability of human language. The larger and more diverse the database, the better the algorithm can understand and generate human language. This learning process results in a "trained model", a separate file where relevant information is stored. However, the creation of these databases often involves copying large amounts of text from various sources, which can potentially infringe on copyright laws as a direct result of its creation process.

The opinion of the Ministry of Justice suggests that the creation of ML databases or datasets could potentially be considered "fair use" under the Copyright Law, falling under the categories of "self-learning" and "research". This interpretation aligns with the spirit of the law, as ML is essentially a form of inductive self-learning. The only difference between human learning and ML is the technical process of learning, which should not be a barrier to the application of "fair use".

The Ministry's opinion further discusses potential market failures and prohibitive transaction costs that could arise in AI enterprises due to copyright issues. The creation of an effective dataset would require negotiating with each copyright owner, a process that could be time-consuming, costly and practically impossible. Delays imposed by any single rightsholder could completely frustrate the entire project, given the competitive constraints and ambitious milestones common in entrepreneurial ventures.

The Ministry's opinion suggests shielding from liability the creation of ML datasets that include vast and

diverse copyrighted works, since, arguably, in such event each individual work included in the dataset holds a relatively immaterial weight in the dataset. The result of this approach is a solution whereby an ex-ante statement is made, declaring that the creation of datasets for ML, in most cases, falls under the fair use doctrine. An ex-ante statement might seem unusual, as fair use decisions are typically made retroactively after the unauthorized use of copyrighted content, but it could be a necessary statement, given the unique challenges posed by ML.

While the Ministry's opinion may mark the direction of the Ministry's approach of the ML/copyright issue, it is important to remember that this is only a guideline, and the final legislation may take a different approach. As such, the opinion serves as an interesting starting point for a broader conversation about the intersection of AI and copyright law, rather than a final word on these matters.

In addition, while the Ministry's opinion provides valuable insights into the implications of copyright law on the creation of machine learning datasets, it stops short of addressing the question of who holds the copyrights to the outputs of the NLP process. This is a significant area of concern that warrants further exploration.

The outputs of NLP raise several intellectual property questions. For instance, who owns the copyright of a text generated by an AI? Is it the developer of the AI, the user who provided the input, or, strangely enough, is it the AI itself (despite the fact that, as of now, AI systems are not recognized as legal entities capable of holding copyrights)? Furthermore, if an AI generates a text that infringes on someone else's copyright, who is liable? These questions become even more complex when considering that AI can generate outputs that were not explicitly programmed by its developers, making it difficult to predict and control the AI's actions and the NLP output and results.

As we delve deeper into the realm of AI and NLP, it is also crucial to address the significant privacy concerns that accompany these technological advancements. These concerns primarily stem from the extensive data collection and processing required for AI and NLP systems to function effectively. The vast amounts of data, often encompassing personal information, raise questions about user awareness and consent. Furthermore, the potential misuse of personal information, through detailed profiling of individuals based on their online behavior, preferences, and interactions, is another area of concern. Adding to these issues is the lack of transparency often associated with AI systems, sometimes referred to as "black boxes" due to their complex and opaque decision-making processes. This lack of transparency can make it difficult for individuals to understand how their data is being used and processed.

The existing privacy laws may not be fully equipped to handle the unique challenges posed by AI and NLP, as they were simply regulated during times in which AI and NLP were almost science fiction. The Privacy Protection Authority (within the Ministry of Justice) recently published an opinion addressing the privacy

concerns associated with "deep fake" technologies, which can create convincingly realistic but entirely fabricated audio and video content. This publication provides much-needed guidance on a particularly controversial aspect of digital technology. Furthermore, the Authority has announced its intention to publish an opinion specifically focused on the privacy aspects of artificial intelligence. This forthcoming opinion is expected to provide further clarity on the unique privacy challenges posed by AI technologies.

This trend towards proactive guidance through publication of opinion letters, rather than reactive legislation, could potentially enable Israel to effectively navigate the complex intersection of technology, privacy and intellectual property rights, which could have significant implications for the country's tech sector.

In conclusion, the landscape of AI regulation in Israel is in a state of flux. The Ministry of Justice's opinion on copyright law and the creation of machine learning datasets is a significant step forward, providing much-needed guidance for the [Israeli tech sector](#). However, there are still many unanswered questions, particularly concerning the ownership and liability of AI-generated content and the privacy implications of AI and NLP. As Israel continues to navigate this complex landscape, it will be crucial to strike a balance between fostering innovation, protecting intellectual property rights, and ensuring privacy. The world will be watching closely as Israel, known for its innovative and leading tech sector, charts its course in this uncharted territory.

## Key Contacts



**Netta Shaked-Stadler**  
Partner