Thesis Project Portfolio

Deep Multimodal Representation Learning to Integrate Natural Language Processing with Genomic Interval Data for Tailored Biomedical Discover

(Technical Report)

Evaluating Social Pharmaceutical Innovation as a Means of Reducing High Therapeutic Costs

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Lilian Jones

Spring, 2024

Department of Biomedical Engineering

Table of Contents

Sociotechnical Synthesis

The current state of pharmaceutical industry operations is unsustainable for both companies and patients. Medication prices are getting exponentially more expensive while returns on R&D are nearing closer and closer to zero, especially in the case of rare disease therapies. The rigidity of the innovation pipeline has continued to drive these costs up. In fact, these therapies can cost over $100,000 per patient per year. It is apparent that there is a critical need to find a new way to balance the needs of pharmaceutical industry stakeholders, regulatory agencies, and patients. Maintaining the economic incentives for innovation and the development of these rare disease treatments while simultaneously ensuring these life-saving drugs are not financially inaccessible to their intended market. Thus, in my STS research I evaluated the prospect of adopting social pharmaceutical innovation (SPIN) as a mode of reframing the prevailing firmly established innovation pathways and overall disrupting the pharmaceutical industry.

I first began by elaborating on what exactly SPIN is by offering a formal definition of this STS concept. With the shift in framework outlined, I then went on to elaborate on the specific practices that, through my research, I assert are the most realistic and effective changes that should be adopted by pharmaceutical companies and enforced legislatively by the US government. Since SPIN is a novel concept and not widely adopted, I conclude this paper by identifying relevant metrics and indices that can be used to quantitatively assess the performance of SPIN strategies. Proposing specific metrics helps holistically understand the ensuing impact and add to the robustness of the framework, thus aiding in the efforts to establish SPIN as the status quo of the pharmaceutical industry. Of the methods investigated through my research, I found that emphasizing collaboration and harmonization between pharmaceutical firms as well as with regulatory bodies, prize fund models, compulsory licensing, and value-based pricing are

methods that show the greatest, immediate promise in improving transparency in the development process and balancing innovation incentives with decreasing the high drug costs patients face.

To then prove these strategies effectiveness, I proposed that these companies regularly conduct patient surveys to gather data on how mortality rates, life expectancy, quality of life, and patient reported outcomes (PROs) have been affected as a result of SPIN. I further elaborate that this data can also be used to calculate quality-adjusted life-years (QALYs) and incremental cost-effectiveness ratio (ICER) to empirically demonstrate the economic impact of SPIN. An increase in the rate of production of patents and the discovery of new molecules by a pharmaceutical company is also a powerful metric in understanding the effect SPIN has on the innovation pipeline of drug R&D. Internal employee surveys can be conducted to also understand how SPIN has affected the morale and efficiency of the company's scientists. In summary, through analyzing the current practices of the pharmaceutical industry and identifying specific changes that can be made through applying the SPIN framework, I have proposed a way that the rigid structure of pharmaceutical innovation can be disrupted. This deviation from traditional practices will incentivize the creation of safe, efficient, and accessible therapies, ultimately catering to unmet needs of rare disease patients and prioritizing social impact over market-driven motives.

The technical portion of my research focused on attempting to address the lack of patient data available to researchers developing rare disease therapies. The inherently limited patient pool creates difficulties in the recruitment of clinical trial participants and gathering necessary data, causing R&D costs to be exceedingly high as compared to therapies treating more conventional aliments. The rationale is to leverage machine learning methods to train a complex neural network on what pre-existing genomic sequencing data is available and generate novel

data to increase the breadth of resources available for scientists to use. My capstone team and I built a generative machine learning model that uses natural language processing (NLP) techniques to take in an input query, such as “pediatric cancer” or “glioblastoma”, and output relevant genomic interval data that can then be used in subsequent in silico bioinformatic analysis. The neural network mathematically transforms the natural language text into vectors of numbers so that the similarities and differences between the words and genomic data can be calculated and understood by the computer. The outputted genomic interval data from the model contains information such as the location of relevant genes, transcription factors, chromatin accessibility and methylation which can be used in biomedical experimentation. By creating this generative tool, less resources and capital will need to be invested in the infrastructure of rare disease clinical trials, reducing the expenses of drug R&D, and thus taking a technical approach to motivate pharmaceutical companies to affordably price their products while maintaining innovation incentives.