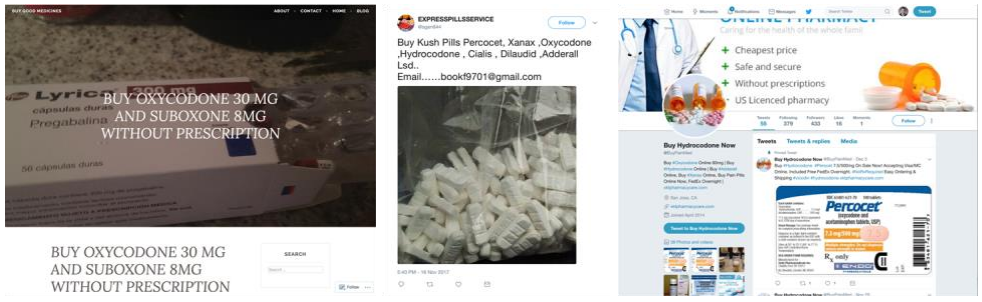
Project Title: Automated Detection, Classification and Reporting of Illicit Online Pharmacies Selling Controlled Substances

Research Idea/Project information:
Illicit online pharmacies that illegally sell prescription opioids direct to consumers via the Internet and social media, often marketing such products as “no prescription needed”, are a clear and documented patient safety and public health risk that is perpetuating substance use disorders (SUD). Despite documented injury and even death to consumers (such as the tragic 2002 death of teenager Ryan Haight who purchased Vicodin online and died from overdose), there are few solutions for federal regulators, law enforcement agencies, and pharmaceutical manufacturers to detect this form of illegal activity for the purposes of regulatory or criminal enforcement. This despite a Federal Law, known as the Ryan Haight Online Pharmacy Consumer Protection Act (RHA) that expressly prohibits the online sale of controlled substances from illegal online vendors. Our proposed SUD Startup Challenge project addresses this gap by accelerating early stage technology we have previously developed using big data and machine learning to accurately detect illicit online pharmacies marketing controlled substances via social media. The system will comprise of automated and real-time data feeds of public APIs from popular social media platforms filtered for key controlled substances brand and INN names that would generate large volumes of unstructured data. This data will then be analyzed using different machine learning algorithms based upon the type and structure of social media data. This includes the use of multiple machine learning protocols to optimize anomaly detection specific to illegal online marketing and sale of prescription opioid drugs as has been used in prior published studies conducted by the team and as presented at the recent HHS Opioid Code-a-Thon where we were selected as a finalist in the “prevention track” (see **Figure 1** for examples). The system will then automate web scraping of any hyperlinks that redirect consumers to illegal online pharmacies included in analyzed content, which would then be content coded to confirm a website's legal status (e.g. illegal or registered online pharmacy, address and IP location of the online pharmacy, etc.) We would then automate a script to report such illegal activity to the U.S. Drug Enforcement Agency (DEA) – the agency charged with administering the RHA-, the U.S. Food and Drug Administration (FDA) Office of Criminal Investigations **(**e.g. DEA via https://apps.deadiversion.usdoj.gov/webforms/jsp/umpire/umpireForm.jsp; and the US FDA via https://www.fda.gov/Safety/ReportaProblem/ucm059315.htm and https://www.accessdata.fda.gov/scripts/email/oc/oci/contact.cfm), and to relevant manufacturers and their security departments. The system would also maintain a historic and up-to-date database of all illegal online pharmacies selling controlled substances as detected by the system and proactively notify any online users that interacted with content that may place them at risk of purchasing a controlled substance online.

Figure 1: Examples of Illicit Marketing and Sale of Controlled Substances Online Detected in HHS Code-a-Thon



Research Team and Technical Competence:
The team is comprised of a multidisciplinary group of researchers with backgrounds in addiction science, public health, computer science, and health technology from the University of California, San Diego. The team lead, Prof. Tim Mackey is an Associate Professor of Anesthesiology and Global Public Health at UC San Diego School of Medicine and is also the Director of Healthcare Research and Policy at UC San Diego Extension. Prof. Mackey has co-authored over 130 publications in various scientific and medical journals and has had his work featured in high-impact journals such as Science, JAMA, Nature Biotechnology, the Lancet, Nature Reviews Clinical Oncology, Clinical Microbiology Reviews, and BMC Medicine. His research and expertise has also been featured in major news outlets such as CNN, NPR, STAT News, and POLITICO Pro. His work focuses on an array of multidisciplinary topics in domestic and global public health and he is an internationally renowned expert on the topic of online drug safety and controlled substances (the focus of this proposal). He also has extensive professional experience including over 10 years experience in the private sector and acting as a consultant for the World Health Organization, the US Department of State, US Department of Justice and others. Janani Kalyanam is a current post-doctoral researcher at UC San Diego and holds a PhD in data science with an emphasis on machine learning. Dr. Kalyanam is a current Postdoctoral Researcher at UCSD and has also worked at the Data Science group in NEC Labs and has published several papers with Prof. Mackey. She has advanced training and experience in "big data" analysis, data mining, and developing machine learning algorithms specifically examining public health issues. She is the recipient of numerous awards including Best Paper Award at the ACM 21st Conference on Knowledge Discovery and Data Mining. As previously mentioned, Prof. Mackey and Dr. Kalyanam led a team that competed in the "prevention track" of the December 2017 HHS Opioid Code-a-Thon hosted by the U.S. Department of Health and Human Services. The team was selected as a finalist during the Opioid Code-a-Thon for its solution that forms the basis for this SUD Startup challenge solution. Further, all members of the research team have collaborated on published research that has established the viability and underlying technology for the proposed project and an MVP. Hence, this uniquely situated interdisciplinary team has the professional, academic, and technology background needed to ensure successful completion of this project and has already been successful in prototyping the solution and having it recognized as both innovative and important in the fight against the opioid epidemic during the Code-a-Thon competition.

Solution Design, Prototyping, and Practical Use and Application:
Our solution is the first to specifically identify the marketing and sale of controlled substances via social media. It will use advanced machine learning algorithms that requires minimal human supervision, one that is unique in its application to addressing SUD and illegal access online. The solution that translates our prior research and Code-a-Thon POC work into an MVP will consist of web-based visual interface that includes: (1) a dashboard of real-time or historical social media data that is collected, filtered and analyzed as associated with illicit online sale of controlled substances that can be queried for any combination of date and name of controlled substance products (data will also include users that have interacted or been exposed to content advertising sale of controlled substances for health promotion purposes); (2) visualization of hyperlinks/URLs that have been coded and classified as illegal online pharmacies accompanied with risk score assessing how likely they are to be an illegal online pharmacy (including information about their IP and registrant address location); and (3) a summary of reports generated and sent to law enforcement and regulators reporting detection of illicit online pharmacies selling controlled substances for further action. The MVP will serve as a platform to demonstrate the viability of the technology and the ability to scale it to larger volumes of data (e.g. data collection and analysis that can be conducted in near real-time and for longer periods of data collection) that can be customized for customer needs (e.g. brand/IP

protection or sale of controlled substances). The MVP will also demonstrate how the solution can be segmented for private (subscription and/or license) or public (educational-free) use. A summary of the POC developed for the Code-a-Thon that will form the basis for the MVP in provided in the **Figure 2** below:

Figure 2: Screenshots of working POC

Methods to Determine Product Need and Target Audience:

Currently, the global trade in fake, counterfeit and falsified medicines is estimated in the billions of dollars, with a high percentage of these sales in fake medicines occurring online. In fact, it is estimated that there are over 35,000 "online pharmacies" in operation, but Internet security firms have also estimated that 96% of these online vendors operate illegally (including "no prescription" online pharmacies and those that sell counterfeit/fake drugs). Regulators, pharmaceutical manufacturers, law enforcement, and customs officials require advanced digital surveillance tools to identify illegal online marketing and sale of prescription drugs, but even more specifically, controlled substances given the current opioid and counterfeit fentanyl crisis being experienced in the United States. However, though a few solutions exist, none are available to monitor illegal online drug sales occurring via the growing array of popular social media sites (Twitter, Facebook, Instagram, etc.) that directly market to and target consumers. Hence, the proposed innovation will address this critical market gap by developing a total end-to-end solution based on advanced machine learning algorithms allowing us to analyze large volumes of unstructured social media data to specifically identify illegal controlled substance online pharmacy sales and to strengthen enforcement of the RHA. The innovation will also act as an important tool for pharmaceutical manufacturers to engage in brand and IP protection and allow law enforcement and regulators to be proactively informed of websites in violation of Federal and state law. Importantly, the innovation combines market opportunities available in the pharmaceutical security sector and burgeoning cybersecurity industry (estimated in the hundreds of billions) as it will combat online pharmacy cybercrime (associated with viruses, phishing, malware, etc.) and also limit financial and consumer fraud.

Assessing the need and market viability for this technology will involve conducting market research, conducting demos of the MVP, and reaching out to known contacts at FDA OCI, DEA, the National Association of Boards of Pharmacy (NABP), pharmaceutical companies, and patient safety organizations, all of whom have an interest in protecting public health and combating the opioid epidemic and can directly benefit from the use of the proposed solution.

Law enforcement agencies can use this technology to identify criminal activity associated with pharmaceutical crime, diversion, and illicit access. Other Federal and state drug regulators, such as the US FDA and State Boards of Pharmacy, will similarly benefit given their regulatory mandates in this area. However, our primary market for this technology will be the pharmaceutical industry and web technology providers that focus on brand/IP protection and online consumer safety. This includes drug manufacturers that spend millions of dollars on pharmaceutical security, but lack solutions addressing the growing and largely unregulated social media landscape and its association with illegal pharmaceutical sales and counterfeiting. Technology providers, such as Google (which was fined $500 million by the US DOJ for illicit online pharmacy ads) and Amazon (which is entering the online pharmacy market), can also utilize our technology for corporate compliance and consumer safety. The public will also be a primary beneficiary as exposure to illegal online pharmacies selling controlled substances will be mitigated.

Reflecting the urgent need for this technology, the development of the machine learning algorithms and general methods underlying our proposed Proof of Concept were initially funded by the Alliance for Safe Online Pharmacies (ASOP), a 501(c)(4) social welfare organization engaged on the issue of illicit online pharmacies. ASOP enjoys the membership of a number of different private sector, professional association, and nonprofit entities, all of which are focused on consumer and online safety. ASOP funded a pilot research project led by the research team that resulted in four published papers in the American Journal of Public Health, the Journal of Medical Internet Research, Addictive Behaviors, and F1000. These initial funding sources acted as the catalyst for the proposed scale up, MVP development, and pursuing the commercial viability of the project proposed.

The original source of funding was focused on generating published research on the viability of the technology to detect illegal online pharmacies selling controlled substances using unstructured and large amounts of social media data. The research team has completed these aims and validated the methodology and potential usability of this technology through several published papers as mentioned previously. However, the original ASOP pilot grant did not include funding to translate this technology into a commercially viable prototype or MVP. During the team's participation in the HHS Opioid Code-a-Thon, a POC was developed to demonstrate the utility, technology linkage and user interface of the project POC. Being selected as a finalist helped confirm that this project has broader viability, hence, funding through the SUD challenge would allow us to fill this critical funding and translation gap in order to scale up the technology by developing a MVP as a software as a service (SaaS) model that could be commercialized via a licensing or subscription model for interested parties. We would also explore a public facing license-free interface/API that could be used for consumer education and awareness. Importantly, funding would allow us to develop a functional MVP that could be demoed for our target audience, many that have already shown interest in the technology described (this includes ongoing discussions with the National Association of Boards of Pharmacy, Pfizer, Merck & Co, and other pharmaceutical manufacturers.) Further, the service could be offered as a more comprehensive monitoring and brand protection solution with additional consulting and customizable services.