Written Testimony of:

Dr. Daniel R Swale
Associate Professor of Insecticide Toxicology

Emerging Pathogens Institute, Department of Entomology and Nematology
University of Florida

Before the:
Committee on Science, Space and Technology

United States House of Representatives

"Innovations in Agrichemicals: AI's Hidden Formula Driving Efficiency"

May 20, 2025

Chairman Franklin, Ranking Member Amo, and Members of the Committee, thank you for the opportunity to testify this morning. I am Daniel Swale, Associate Professor of Insecticide Toxicology and Neurophysiology in the Emerging Pathogens Institute and Department of Entomology and Nematology at the University of Florida. I appreciate the House Committee on Science, Space, and Technology holding this important hearing and I look forward to explaining my work in the field of agrochemical discovery, the incorporation of artificial intelligence within agriculture at the University of Florida, and providing an overview of my personal views regarding the opportunities and challenges of artificial intelligence within the field of agrochemical discovery and the early phases of the developmental pipeline.

I have more than 15 years of experience within the field of agrochemical discovery and development that includes the discovery of novel chemical technologies to control insect and fungal pests of row- and specialty- crops, insect and arachnid pests of livestock, as well as insect and arachnid vectors of human pathogens (e.g. mosquitoes and ticks). These pests directly lead to billions of dollars lost in economic revenue in the agricultural sector and due to a multitude of reasons ranging from pesticide resistance to increasingly stringent environmental regulations, the ability to control these pests has become severely limited. While new technologies, such as genetic engineering, have potential to improve agricultural sustainability in the distant future, the use of chemical pesticides is a proven approach for rapid reduction of pest populations and disease. Considering this, the overarching goal of my research program is to utilize new technologies, such as artificial intelligence, to discover novel pesticides that are safe, effective, and marketable to ultimately support sustainability of our agricultural sector.

Today, my testimony focuses on: (1) the ongoing investment in AI-driven technologies at the University of Florida and the utilization of AI in agrochemical discovery in my research program; (2) my personal viewpoints regarding opportunities and challenges of incorporating AI in agrochemical discovery and development; and (3) the steps required to fully maximize the potential of AI in agrochemical discovery to ensure a return on investments.

There are a significant number of opportunities for utilizing AI/ML technologies to improve agrochemical discovery and development, but there are also many barriers that must be addressed prior to being able to maximize the potential of AI/ML to agrochemical discovery and development. The gaps in appropriate and robust datasets, such as physiochemical properties or delivery modalities, do not exist for the many different pest complexes and this gap must be filled before we can meaningfully expect to maximize the power of AI in agrochemical discovery**.** Further, utilizing AI platforms to move chemical discovery from the private to the public (e.g. universities) sector will allow data (e.g. algorithms, inputs, etc) to be public and promote innovation and development at the national and global levels.

Ultimately, my goal is to provide a general, non-technical outline that can be used to promote leadership and global competitiveness by the U.S. in agrochemical innovation with an emphasis on agrochemical discovery. I look forward to conversations with the esteemed members of this subcommittee regarding this important issue, and I am happy to elaborate further on any areas of inquiry regarding my testimonies.

Background Justifying the Need for Investments in Agrochemical Discovery

As the population of the world rapidly increases with estimates of the global population reaching 10 billion by 2050 – which is 2-3 billion more mouths to feed compared to today – finding answers to the question of how we grow more food, in less space, with an increasing prevalence of agricultural pests becomes even more pressing.

The United States has historically been a global leader and has dominated the landscape pertaining to innovation and development of novel agricultural technologies and in particular, agrochemicals. We no longer are the global leader in agrochemical innovation or agrochemical development with Japan exceeding the U.S. in numbers of first-in-class compounds by over 2-fold and China near equal to the U.S. with total numbers of new insecticides. And even more concerning, I personally believe the gap between other countries and the United States with regards to agrochemical innovation is expanding rapidly.

So how and where do we employ AI technology to improve U.S. agriculture? I believe one of the largest gaps ensuring economic viability and sustainability of our agricultural industry is due to the reduced number of chemical technologies available for rapid and effective control of pest populations. Thus, I believe the largest gains in technology innovation and most significant contributions to be made for agricultural sustainability are within the discovery phase of agrochemical developmental pipeline. In other words, how can we employ AI technologies to increase opportunities to discover new and effective chemicals that can be quickly registered and delivered to farmers?

The current challenge with commercialization of novel pesticides is the tremendous cost and time from discovery to commercialization. Current estimates are a novel mechanism pesticide ranges from $500 million to $1 billion dollars and takes approximately 12 years from discovery to the farmers hands. This high expense and prolonged timeline has functionally reduced the numbers of agrochemicals being commercialized and led to repurposing of existing materials versus innovating and developing unimagined chemicals. Some of these costs and developmental timeline are unavoidable, but I do believe thoughtful and intentional integration of AI into this developmental pipeline can be used to reduce costs and time of development to promote innovation in discovery.

The goal of my testimonies is to explain the opportunities of artificial intelligence for the field of agrochemical discovery and challenges that must be addressed for us to harness these advancements in technology, apply them to the agricultural sector with a particular focus on chemical discovery, and restore the United States as the leading innovators of technology and products across the world.

AI Infrastructure for Chemical Discovery and Agriculture at the University of Florida

The Emerging Pathogens Institute (EPI) at the University of Florida (UF)) is a uniquely positioned inter-disciplinary institute to establish a program that will enhance data science- and artificial intelligence (AI)-driven discovery for agronomical and agrochemical research that can lead to agricultural sustainability and the discovery of new agrochemicals. Through a collaborative and translational research hub via transdisciplinary "*Team Science*", UF EPI researchers currently aim to (1) develop advanced computational models, methods, and tools, leveraging heterogenous data resources, (2) design, implement, and evaluate data science and AI-based interventions to combat plant pathogens, and (3) discover novel agrochemicals to kill pests that reduce the economic sustainability of Florida and U.S. agriculture.

In addition to ongoing work in the agricultural space at the UF Emerging Pathogens Institute, it was recently announced an Artificial Intelligence center will be coming to the UF Gulf Coast Research and Education Center. This new center will be a 19,000 square foot AI hub that will focus on utilizing the power of AI to develop novel technology to improve U.S. Agriculture through improved plant breeding as well as other critical aspects of agricultural sustainability. I believe the University of Florida is well positioned to promote national and global innovation and technology advancement throughout the agrochemical development pipeline due to the world-class expertise within the AI space combined with the significant infrastructure in AI.

Advancing Innovation and Technology Development Through AI-based Discovery in Public Sectors of Research

Using the classical approach for agrochemical discovery, where a chemical is discovered through manual testing of hundreds of thousands of compounds against a single pest, there is a need for large investments in infrastructure, money, and personnel that is not currently feasible in academia given limited funding from federal agencies for this activity. Thus, discovery and early-phase development of agrochemicals has been restricted to the private sector. While private industry is certainly capable of exploiting artificial intelligence to develop new technologies within the agrochemical sector, the methods and details of "how" AI was used to develop new agrochemical technology is unlikely to become widespread due to intellectual property restrictions. With the correct implementation, investing in AI-driven discovery at universities can bridge many existing gaps and stimulate innovation at the state, federal, and global levels.

I believe the advent of AI enables public entities, such as universities, to play a key role in the discovery and development pipeline of agrochemicals and potentially, play a leading role due to their ability to test fundamental questions oftentimes ignored by industry. Thus, movement of agrochemical discovery currently restricted in the private sector to universities will allow data (e.g. algorithms, inputs, etc.) to be public and promote innovation and development at the national and global levels. The University of Florida is well-suited to perform agrochemical discovery through world-class expertise in agrochemicals, existing infrastructure in AI, superior supercomputers with large data storage capabilities, and the development of an AI hub for agriculture at the UF Gulf Coast Research and Education Center. **I encourage Congress to consider investments into academic institutions that can participate in the early phases of agrochemical discovery and development, which will expand opportunities for innovation of novel technologies and promote public-private partnerships at the state and national levels.**

Ongoing Research in the Swale Research Laboratory at the University of Florida

The Swale research lab focuses on the development of novel chemical technologies to address the economic and human health concerns resulting from arthropod, fungal, and bacterial pests. Our current research lies at the interface of physiology, toxicology, and molecular genetics to discover and develop novel pesticides, provide knowledge on the modes of action of pesticides, and characterize resistance to insecticidal chemistries. More specifically, the Swale Lab aims to define the physiological role and toxicological relevance of ion channels and ion transporters that are understudied to guide the discovery and development of novel chemical pesticides.

While the Swale lab employs classical approaches for the discovery of new chemical pesticides, our group is also investing effort into utilization of Artificial Intelligence systems to promote discovery of novel active ingredients. These projects are briefly summarized below.

1) Using innovative AI technology to bring natural-product based compound discovery to agriculture

Nature itself provides many successful and safe pesticides, but unlike DNA and proteins, the chemical world remains vastly unexplored—more than 99% of chemical structures in nature are still unknown, leaving immense untapped potential for innovation. Understanding chemistry allows us to decipher the "why" behind biology and in the context of agrochemicals, identify and develop natural pesticides that can promote sustainability of the U.S. agricultural sector.

Despite nearly 99% of the chemical structures in nature remaining unknown, approximately 70% of all FDA-approved small molecule drugs are derived from or inspired by natural products and approximately 50% of agrochemicals have been or could have been derived from natural sources. Why? Because natural chemicals are inherently optimized in living organisms, making them highly effective and compatible with biological systems. Due to this, natural products (NPs) have been a particular focus in the discovery of new agrochemicals and testing of NPs for activity to pests of agriculture represent one of the most common approaches employed by research and development (R&D)-based companies involved in agrochemical discovery. NPs in the context of chemical discovery refer to chemical compounds or substances that are produced by living organisms including plants, animals, or microorganisms.

Natural products are the result of millions of years of evolution and thus, plants have evolved sophisticated chemical defenses to protect themselves from pests and pathogens, as they cannot simply move away from threats. Unlocking these natural pesticides relies heavily on our ability to understand and replicate the complex chemistry these plants have developed. **The barrier has been the absence of technology capable of comprehensively understanding chemical structures and interactions at scale**—essentially, a "chemical sequencer." But that capability to interpret the complex molecular language of nature is now emerging. Today, leveraging advanced artificial intelligence, companies like Enveda (Boulder, CO), are at the forefront of this transformation as they have developed the capability to "read" and predict chemical structures and biological activities with unprecedented speed and accuracy. Their approach is similar to how AI models like ChatGPT have mastered human language.

Through collaboration between the Swale Lab and Enveda (Boulder, CO), which is an AI-focused company specializing in discovery of NP-based human drugs, we are applying this

innovative AI technology to address the historical gaps in knowledge or technology to discover and develop natural agrochemicals for pest control. To do this, we've assembled the world's largest database of plant chemistry to answer where to find these natural compounds. Second, we separate the plant into hundreds of components, or fractions and then apply a large language model-based ML algorithms to predict the chemical structures and properties of all the compounds, which are in the thousands, that exist in the plant. Next, we test the identified and extracted compounds on a pest of interest to quantify biological activity, or toxicity, of the compound to the pest. When taken together, this approach has bridged multiple barriers that prevented advancement in the 1990s and early 2000s and we are optimistic this approach will lead to the discovery of novel active ingredients for control of insect, fungal, or bacterial pests of agriculture. **I encourage congressional support of public-private partnerships to promote growth of AI, natural products, and protection of our food supply, which will drive continued innovation and advancement of the U.S. on a global scale.**

2) Development of novel chemical insecticides to protect the citrus industry

The citrus industry, which contributes nearly $3 billion USD to the U.S. economy, is in dire need for novel approaches to control an insect pest that transmits a bacterial pathogen to the citrus tree, which is the causative agent of citrus greening disease that has led to near collapse of the Florida and U.S. citrus industry. One approach to restore citrus sustainability is the development of novel insecticides to control the insect pest and thus, the Swale Lab has developed a dual-mechanism insecticide that is acutely toxic to the insect pest while simultaneously preventing feeding at concentrations that are equal to commercialized insecticides. This dual approach functions to reduce the insect population and prevent pathogen transmission from the insect to the tree. However, the extraordinarily high financial costs of development and the prolonged registration pipeline for synthetic insecticides will likely challenge the progression of these newly developed to move from "bench-to-field" and help restore the citrus industry.

To overcome this challenge, the Swale lab has focused on identifying compounds from the natural world that structurally resemble the lead synthetic insecticides. The costs and timeline for development of natural chemistry is significantly lower when compared to synthetic pesticides, which makes this a promising approach for agrochemical development, but significant challenges exist for identifying natural analogs. For instance, where do we look for natural analogs, how do we predict structural similarities between synthetic and natural compounds, and how do you minimize "false-positives."

To remedy this, the Swale Lab has employed Artificial Intelligence and Machine Learning platforms to identify plant families and/or known natural products that contain chemicals that are structurally related to our lead (synthetic) compounds. Computational tools for focused compound libraries require scoring systems to evaluate natural product-likeness within a chemical space and using this approach, we have identified multiple natural sources that contain chemicals with structural similarity ranging between 80-90% when compared to the synthetic compounds developed by our lab. These natural sources have been obtained, and chemical isolation protocols are ongoing to test the biological activity of these natural compounds. Downstream work aims to employ machine learning to refine the natural product-likeness scoring system for these series of compounds to enable efficient analysis of larger natural product compounds for insecticide-like properties, metabolite-likeness, and lead-likeness.

The incorporation of AI and ML approaches to discover natural analogs of synthetic leads is a significant advancement for the field of agrochemicals because structural likeness between synthetic and natural compounds is oftentimes not obvious. Natural products display unique structural features, such as diverse shapes, no fluorine molecules, large and complex ring structures, and complicated structures. The ability to employ AI for identification of novel natural analogs to known synthetic pesticides has the potential to provide new active ingredients to the market that are significantly cheaper and quicker to commercialize.

Despite preliminary success, there are limitations of current AI methods for natural agrochemical discovery that should be addressed to ensure maximum utility of the technology. First, limited data availability is a major barrier for exploiting AI for discovery of de novo agrochemicals from natural origins or identifying analogs of known active compounds. Appropriate and strong inputs are essential for obtaining accurate and high-quality outputs from AI or deep learning approaches, yet these inputs are oftentimes missing from natural products. Natural product databases oftentimes lack comprehensive data on chemical structures, biological activity, and chemical properties that complicates the training of accurate AI models. Increased robustness of natural product databases would tremendously enhance the utility of AI in natural product agrochemical discovery.

Second, the complexity of natural product chemistry makes it challenging for AI algorithms to predict activity in agricultural pests due to a lack of knowledge of "what makes a pesticide active." Increased robustness of chemical properties that drive activity to a particular pest, such as the Asian citrus psyllid which drives the citrus greening crisis, will assist in identifying natural compounds with biological activity towards the pest of interest. Importantly, the Swale Lab and the University of Florida are poised to address these fundamental, but critical questions to improve the utility of AI platforms for discovery of novel agrochemicals by taking advantage of existing AI infrastructure at UF and development of an agrochemical discovery component of the new AI hub for agriculture at UF Gulf Coast Research Station. Further, public-private partnerships similar to the collaboration with Enveda will aid in mitigating some of the stated barriers with natural product discovery.

3) Utilizing Artificial Intelligence to enhance classical *in silico* modeling for agrochemical discovery

Computer aided chemical discovery has been utilized across pharmaceutical and agrochemical platforms for decades, but has remained a modest component of drug or agrochemical discovery process. *In silico* screening for chemical leads is a process where computers are used to virtually assess the affinity of chemical structures against a known protein target through modeling interactions between chemical molecules and the biological target. Historically, this was performed with homology models, which provide a general estimate of the protein structure but is unrefined and lacks detail, which resulted in identification of compounds with modest activity. Recently, *in silico* screening has become more productive with advancements in defining the structure of protein targets via cry-electron microscopy, the expansion of virtual chemical repositories, advancements in computational methods to predict exponentially more ligand-protein interactions, and increased computer power. The Swale Lab aimed to explore the potential of utilizing artificial intelligence and Machine Learning approaches to identify unrealized chemical scaffolds that have activity to potassium ion channels in the insect salivary gland.

Aphids are an insect pest that feeds on thousands of different plants and are responsible for $1+ billion dollars lost due to their feeding on agricultural crops, including citrus, strawberry, soybean, cotton, and corn. The economic relevance of aphid feeding justified our efforts to identify novel approaches to prevent aphid feeding. Our previous work has shown that potassium ion channels are critical for proper functioning of the aphid salivary gland and inhibition of these ion channels through chemical inhibitors prevents salivation and inhibits feeding biology of aphids. While we had active chemicals that prevented aphid feeding on plants, we aimed to expand the diversity of chemical structures available through in silico screening methods. However, we were met with challenges due to the lack of known structure of aphid potassium channels and we had minimal ligand training datasets that are required for *in silico* screening.

To remedy this, we collaborated with a start-up company to test the utility of an innovative machine learning approach to discovery high affinity ligands with only the protein sequence and does not require a crystal structure, utilizes ligand training data, and has high computational efficiency enabling the screening of trillions of compounds. Using this method, we identified approximately 100 high-affinity compounds and a portion of those compounds prevented aphid feeding after treatment of plants. However, while the binding affinity is high, the concentrations required to inhibit feeding in the living aphid is 1-2 orders of magnitude higher than currently commercialized aphicides. This loss of potency and translation between virtual and "real" efficacy is common within agrochemicals and the parameters driving this gap in translation must be defined to improve the efficacy of AI-driven discovery for agrochemicals.

The discovery of novel and highly potent chemical ligands based on only a protein structure through AI and ML technologies is a tremendous advancement over the previous approaches for performing virtual screening to discover agrochemicals. However, while we have high optimism that this approach can be used to enhance agrochemical discovery, challenges with this approach still exist and these gaps must be addressed prior to maximizing the potential of this approach.

First, there are tremendous challenges in translating compounds identified via virtual (*in silico*) screening platforms to realized (*in vivo*) activity in an insect pest due to challenges of delivery and pharmacokinetics. To remedy this, knowledge regarding chemical properties driving insect specific activity is required. Secondly, the physical properties of human pharmaceuticals and agrochemicals is starkly different, but unfortunately nearly all the virtual chemical repositories contain chemicals that have pharmaceutical-like properties. Thus, the outputs from AI-drive discovery approaches are likely to identify compounds that are great ligands of proteins but have poor translation to the living insect.

A focus on bolstering the availability of virtual or physical chemical repositories that contain agrochemical-like properties is needed. These fundamental questions can be addressed and answered by academic scientists and will promote advancements to the delivery of AI-generated agrochemicals, yet opportunities to obtain federal funding to address these fundamental questions are currently poor.

Opportunities of AI in Agrochemical Discovery

Note: My research career and expertise is focused on the early discovery and chemical development phase of the agrochemical development pipeline and is not focused on the field efficacy, environmental fate, or other regulatory studies performed in the late stages of development. Thus, I focus this written testimony on these aspects of agrochemical discovery and do not focus on the components to safety testing, commercialization, AI-contributions to pest management, or other areas AI has contributed to agrochemicals.

The tasks of finding successful new drugs is daunting and predominantly the most difficult part of drug development. This is caused by the vast size of what is known as chemical space, which is estimated to be in the order of 10^{60} molecules and thus, technologies that assist with the discovery of high-quality chemical leads would be an invaluable contribution to the field of agrochemicals. Although the utility of AI or machine learning approaches in agrochemical discovery remain rather limited, there is optimism based on the relative growth within the human pharmaceutical industry.

Agrochemical discovery is unique from pharmaceutical discovery in that the target for drug binding is less relevant because the desired phenotype from agrochemicals is death of the pest or pathogen whereas in pharmaceuticals, it is modulation of a specific system linked to a disease. Thus, pharmaceuticals identify the target linked to disease and perform target-based discovery for new chemicals, yet this approach has never led to the commercialization of an insecticide and is also poorly effective for herbicides or fungicides. Thus, agrochemical discovery relies on *in vivo* screening of large chemical libraries to give you the phenotype that you are looking for, oftentimes death, of that organism and the target is unknown. Understanding the chemical parameters that facilitates toxicity to a pest organism is critical to improve translation from AI-discovered chemical leads to organismal efficacy.

Furthermore, the chemical properties of human pharmaceuticals are different than the properties of most agrochemicals, which is a critical point to consider because most publicly available chemical libraries, virtual or physical, contain pharmaceutical-like and not agrochemical-like compounds. Thus, the virtual or physical screening of these libraries has a lower likelihood to discover novel agrochemicals effective at the organismal level simply. Development of publicly available virtual or physical repositories that contain agrochemical-like compounds versus pharmaceutical-like compounds will improve the outputs of AI-generated discovery.

In agrochemical development, there are at least 3 general tiers that must occur in sequence for product registration: discovery, development and lead optimization, and safety testing for registration pipelines. In addition, it is necessary to mention the potential of AI to reduce cost and time requirements for agrochemical development. It is important to note that this process is different and thus the contributions of AI for this general process is likely to be different for insecticides, fungicides, and herbicides and starkly different when compared to pharmaceuticals.

Areas Where AI and ML Can Contribute to Agrochemical Phenotypic Screening

Discovery: The utility of AI approaches in agrochemical discovery remain rather limited, but optimism exists due to the success of AI in pharmaceutical discovery. I describe 6 areas where AI could be used for improving agrochemical discovery: 1) Chemical Library Development, 2) identification of new leads, 3) discovery of natural compounds, 4) drug repurposing potential, 5) compound design and synthesis schemes, and 6) mode of action identification. The Swale Lab and the University of Florida are poised to address each of these aspects of within the discovery phase for agrochemical development by leveraging the existing AI infrastructure at the University and the investments in AI centers focusing on agriculture.

Additionally, the focal point of "discovery" has been on the identification of active ingredients for agrochemicals. Yet, a much needed contribution to the field of agrochemicals is the development of new components of the formulated product to improve delivery of the active pesticide to the pest of interest. Similary, efforts to identify novel synergists that mitigate pest-mediated metabolism of the compounds is critical for sustainability of the newly discovered active pesticide. Knowledge or efforts in these fields have remained limited.

Development and Lead optimization: The most used area within the present day for AI and machine learning for agrochemical discovery and developmental pipeline is for lead optimization. After the discovery of a novel active ingredient, the next phase in the pipeline towards commercialization is "lead optimization and development" where the newly discovered chemical is structurally optimized to improve biological activity, improve selectivity profile away from mammals and beneficial animals, and understand metabolism of the compound in the pest and the environment. Artificial intelligence and deep learning algorithms are likely to bolster the number and quality of these structural analogs with improved biological properties as well as provide optimal schemes for synthesizing the compounds.

Safety Testing and Registration: Three of the primary parameters are assessed early in development: impacts to bees, aquatic toxicology, and mammalian toxicology. Early phase AI algorithms exist for these toxicological assessments but oftentimes are restricted to predicting toxicity versus sub-lethal effects that are critical for utility in the market. For example, the neonicotinoids are a class of insecticides that are not acutely toxic to bees and are one of the most commonly used classes across the world, yet were recently banned by the EU due to potential impacts to bee behavior. Development of AI algorithms that can reduce the expensive and prolonged studies for animal testing, environmental toxicology, and determine sub-lethal effects to pollinators would dramatically improve the agrochemical development pipeline through a likely reduction of costs and time to development.

Reduction of costs and time for commercialization of new, unimagined agrochemicals: The incorporation of ML technologies to quickly assess likelihood of progressing past the regulatory "red tape" will save costs and time for development. Similarly, many pesticide candidates cannot be registered because of their high toxicity to mammals or non-target organisms. Therefore, it is extremely important to evaluate the toxicological profiles early in the developmental process, yet due to time and cost effectiveness and other reasons, toxicity evaluation tests still cannot meet the needs of lead candidate discovery and optimization.

In silico prediction of toxicological profiles has become a viable alternative and incorporation of AI/ML technologies can further enhance the quality of predictive outputs. It is necessary to understand the inputs needed to accurately predict the activity to non-target organisms, environmental persistence, etc. and thus, efforts should be placed into bolstering knowledge within this area to improve the depth and quality of AI inputs.

Overview of Challenges and Outlook for AI in Agrochemical Discovery

AI allows enhanced rapid and continually more refined *in silico* screening of millions of compounds, which should reduce the number of compounds needed to be synthesized for *in vitro* or small-scale *in vivo* assays. This will directly benefit lead optimization of existing compounds we know to be toxic to a pest, and it should provide quicker and better leads in early discovery efforts. However, good inhibitors at the molecular level rarely, if ever, translate into efficacious agrochemicals at the level of the field, which is contrary to the successful *in vitro* discovery approach of human pharmaceuticals.

Considering this, the vast number of deep networks required for the outputs of Machine Learning to be useful is currently a major barrier due to the lack of knowledge surrounding the parameters facilitating biological activity in a particular pest. For instance, the learning algorithms need to consider the physiochemical parameters, metabolic profiles of dozens of organisms, biological fate predictions in multiple environments, pest specific biology (e.g. sucking versus chewing insect pest), cost-efficient synthetic schemes, and many other factors. **While knowledge of these inputs isn't the challenge, the challenge of implementing AI for agrochemical discovery lies around the lack of robust datasets for these, and the many other parameters, do not exist for agrochemicals. Filling these gaps in knowledge and gaps in technology represent a requirement for successful implementation of AI in agrochemical discovery.**

While it is likely that AI can assist in chemical library construction or *in silico* screening for hit identification, trying to refine molecular parameters in a general way and relying on these rigid rules to identify chemicals for testing will overlook outstanding pesticides. For example, the glyphosate structure would clearly be discarded based on well established "requirements" needed to have biological activity in plant systems. The fact that we would miss glyphosate in an *in silico* AI-driven screen due to the need for inputs based on physiochemical parameters known to be active in weed systems raises critical questions pertaining to how we develop algorithms for AI/ML to consider the many "outliers" that exist within the agrochemical marketplace.

Challenges of Incorporating AI into Agrochemical Discovery

AI has gained significant interest due to the ability of these technologies, such as deep learning or deep neural networks, to perform equal or better than humans at intellectual games, drive the development of autonomous vehicles, among many other examples. This progress has led some to predict a new wave of technological change that will have generational impacts, which in some fields, is likely true. However, for agrochemicals, applying the tools is not as easy as it may seem and applying them correctly is even more challenging due to significant gaps in knowledge and infrastructure. Further, it is imperative that those groups within the field of agrochemical discovery not rush the implementation of new technologies as history has shown regression periods due to gaps in knowledge or technology that allows for adequate uses of the new approach.

1) Requirement for large, high-quality datasets: Machine learning approaches require a vast amount of high-quality agrochemical data sets without missing values or poor annotation across a vast array of agriculturally relevant organisms. This does not currently exist and this lack of resources will remain a challenge to several AI driven processes, such as knowledge extraction and predictive modeling.

2) Expert implementation in the agrochemical discovery pipeline and AI inputs: This is critical because certain scientific questions require more than basic understandings of chemical principles given the inherent complexity and vast differences across agrochemical disciplines. Thus, these programs must be led and decisions made by classically trained agrochemical toxicologists and not by computer scientists or geneticists with a relatively poor knowledge of chemical design and chemical-protein interactions from an agrochemical perspective.

3) Requirement for determining single pest of interest prior to initiation of AI driven discovery: Currently, agrochemical industries perform phenotypic (e.g. toxicity) screens simultaneously against 2-3 different insect pests, 2-3 fungal pests, and 2-3 weed pests to provide opportunities for serendipitous discovery. This is likely not possible with the use of AI because inputs into the ML algorithms must be specific for the pest of interest. For instance, the chemical properties to deliver insecticides to sap feeding insects versus chewing pests is different and this must be considered during the inputs for AI. Therefore, we must define and consolidate high-quality inputs for each individual pest that could lead to increases in the cost of discovery, at least in the short term, as many studies and baselines must be established to ensure high-quality inputs are used for AI/ML discovery.

4) Gaps in available protein databases for agriculturally relevant pests: The study of gene or protein mutations in resistant pests can be investigated to identify the underlying mechanism and suggest actions for the design of more potent agrochemicals or propose new modes of action. Validation of the proposed hypothesis can only be achieved if protein structures are available for large and diverse sets of agriculturally relevant species (e.g., pests, pollinators). This is, however, a bottleneck in AgChem research that impedes the discovery of novel protein targets and modes of action, as well as the generation and optimization of molecules.

Cautionary Tale: Innovation Paradoxes

The fields of insecticide science and human pharmaceuticals mirror each other, which should allow researchers in the field of insecticide science to heed lessons learned from human pharmaceutical development. The human genome was released in 2000 and was considered the "golden ticket" for the development of new medicines as it was predicted that employing genomic approaches would lead to increased identification of drug targets, increased rate of discovery, reduced cost of production, and lead to the development of safer drug. The piqued enthusiasm and high promise for the development of new targets that was evidenced by the 4-fold increase in the NASDAQ Biotech Index but, interestingly, the release of the human genome had an opposite impact on the pharmaceutical field than predicted.

The Human Genome Project has led to immense advancements for the field of human biology, but it also led to a 50% reduction in drugs developed and increased the cost of discovery by 200% despite the predictions it would have the opposite effect. A retrospective analysis of the

human pharmaceutical industry shows that most new drug applications developed in the 1990s were through competitor-inspired approaches, which was highly productive with 30-50 new drugs and biologics produced each year in the late 1990s. Interestingly, a dramatic decrease in NDAs, and thus research productivity, was seen in the decade following the release of the human genome (2000-2010) with the annual average new drug applications falling from 39 in the late 1990s to 26 per year during the next decade.

In addition to the decline of R&D productivity, the cost of new drugs increased at an annual rate of 10% per year after the release of the human genome and resulted in a significant economic strain for the industry. It is likely the increased average cost required for approved drugs is partly attributed to the influx of unproven targets that led to significant investments in targets that had poor therapeutic potential, among other reasons.

It is critical to remember this reduced productivity in human pharmaceuticals after the introduction of new technology as it provides a strong correlation to ongoing interests in chemical discovery using artificial intelligence. **It is imperative to have the needed framework in place (e.g. agrochemical datasets needed for inputs) prior to upending the successful agrochemical industry in favor of an unproven, albeit promising, technology.**

Outlook for AI Uses in Agrochemical Discovery

The current situation with AI in pesticide discovery is similar to where the field of combinatorial chemistry was a few decades ago. It was thought the technology was going to revolutionize pesticide discovery, but it didn't. However, it did allow faster production of some chemical analogs for screening once an active compound was identified and thus, had significant contributions to lead optimization. That being said, the implications for AI go far beyond combinatorial chemistry, and I am sure that it will have more impact if time and resources are given to establish the needed inputs for AI-directives in agrochemical discovery.

There are a significant number of opportunities for utilizing AI/ML technologies to improve agrochemical discovery and development, but there are also a large number of barriers that must be addressed prior to being able to maximize the potential of AI/ML to agrochemical discovery and development. For instance, the gaps in appropriate and robust datasets, such as physiochemical properties or delivery modalities, do not exist for the many different pest complexes and this gap must be filled to maximize the power of AI in agrochemical discovery.

Suggestions for a Path Forward

Significant emphasis should be placed on building of high-quality, robust, agriculturally specific inputs that can be used for AI/ML because the majority of inputs we currently have publicly available from a chemical perspective consists of antiquated information (i.e. 1950s datasets), small datasets for individual pest complexes and few chemical scaffolds, or based on knowledge gleaned from human pharmaceuticals that can be dramatically different when compared to agrochemicals. A barrier to obtaining these inputs is intellectual property protection from agrochemical industries and a reluctance to share chemical datasets. **Investments in public entities, such as universities, can begin to bridge this barrier of restricted datasets and proprietary knowledge around the inputs for successful AI discovery platforms.**