

## **Statement of Sean McClain, Founder & CEO of Absci**

### **Senate AI Insight Forum: Innovation**

**October 24, 2023**

Thank you, Leader Schumer, Senators Rounds, Heinrich, and Young for convening the Senate AI Insights Forums, and for inviting me and Absci to be a part of this important national dialogue.

We are all here today because AI is revolutionizing and has already revolutionized the various industries and sectors we represent. We can all recognize that there is incredible innovation to harness, ingenuity to nurture, and safeguards to promote as we move forward in this AI revolution.

Much of the public dialogue around AI thus far has been around the transformative AI technology of now household names such as ChatGPT and DALL-E. There are other important applications of AI however- one of these is AI in biology. From biomanufacturing and clinical trials to diagnostics and drug discovery, generative AI represents a huge opportunity for society writ large. The promise that AI holds in health care, as well as safeguarding against risks such as bioterrorism is less often discussed but requires timely discussion by policymakers.

Today, I want to share some first-hand insights on the opportunities and areas of concern when it comes to AI in biology, as well as offer some considerations for policy makers to maintain US Leadership in AI, ensure our field innovates responsibly, and foster the most societal benefit from the transformative potential of this field.

To do that, I'll share my perspective from Absci, a generative AI drug discovery company, and provide some context from the intersection of AI, biotech, and the life sciences industry.

#### **Absci is Pioneering AI Innovation in Biology**

I founded Absci a little more than a decade ago with the aim to decrease the cost of drugs for patients. In that process, we built a scalable platform that generates high-throughput and high-quality biological data at previously unimaginable scale. This would prove to be the key to unlocking the potential of generative AI in Biology.

In the simplest terms, Absci is the ChatGPT of biologics drug discovery. But instead of asking AI to generate text in response to a prompt, we ask it to generate antibody drug candidates in response to a disease target.

Where it typically takes a decade and over a billion dollars to bring one new drug to patients, generative AI – combined with our ultra-high-throughput screening platform – enables us to create and validate millions of antibody candidates in a matter of weeks, which may shorten the timeline to bring drugs into the clinic. That has the potential to transform the drug industry and help us achieve our mission to bring better biologics to patients faster.

Today, Absci can screen billions of cells per week, allowing us to go from data to train, AI to create and wet-lab to validate in as little as six weeks. Cycles can be run in parallel, allowing us for the first time to generate the high-throughput and high-quality biological data needed to scale AI, and unlock the true potential of generative AI in biology.

## **Data is King**

While AI has already disrupted various industries, biologics drug discovery has largely been undisrupted until now. That's because of the scarcity and difficulty of scaling of high-quality biological data. In the context of antibody drugs, Absci has solved the problem of scalable biological data, which allows us to go from finding needles in a haystack, to creating the needle. This changes the paradigm from drug discovery to drug creation.

But to have an effective AI drug creation platform, you need massive amounts of high-quality biological training data. In fact, I would argue that training data is more important than AI models.

There are several important public data sources that are vital to U.S. innovation in biotech and beyond. These open-source datasets are important resources for drug discovery – we should support them in every way we can through policies that encourage collaborative scientific advancement while instituting robust safeguards against potential bioterrorism and unintentional accidents.

But at Absci, we realize that open-source data alone is insufficient to scale for AI. That is why we build on public data sources by adding our own experimental wet lab data. Our proprietary synthetic biology platform generates the massive amounts of biological data needed to enable the predictive power of machine learning. Then, we use that data to design and run experiments *in silico* (on the computer rather than at a lab bench). This dramatically speeds up the discovery process at a scale and pace that could never be found using standard methods. Supporting the development of a vibrant biotechnology community, with pooled proprietary datasets, will help set the U.S. apart when it comes to scaling biological data for AI use.

## **State-of-the-art AI Innovations**

The promise of AI to fundamentally disrupt the traditional drug discovery paradigm isn't the future, it is happening here and now through the work of Absci. Our generative AI platform has already achieved several "firsts" in AI for biology. Here are some specific examples of how Absci has applied AI innovation to advance the state of the art in our industry:

We were among the first to create a natural language model capable of predicting protein function (Schwartz et al. 2018) and among the first to create a natural language model for antibodies (Bachas et al. 2022), demonstrating that the binding-affinity of an antibody to its antigen is a trait that can be learned and predicted by AI. Absci has also successfully applied its platform to infectious disease, rapidly designing a single therapeutic antibody for multiple SARS-CoV-2 variants.

- More recently, Absci was the first to apply generative AI to design a novel antibody from scratch using zero-shot generative AI (Shanehsazzadeh et al. 2023).
- In collaboration with Caltech, Absci used patient data to identify 12 neutralizing antibodies to SARS-CoV-2 (COVID-19) with equal efficiency to the commercially distributed Regeneron antibody. Additionally, Absci recently began a collaboration with Caltech, supported by the Bill & Melinda Gates Foundation, to develop neutralizing HIV-antibodies.

More broadly, Absci's AI platform has been and is being commercially deployed to design and optimize therapeutic antibodies to a variety of disease antigens with large partners such as Merck, and NVIDIA among others.

### **Policy Considerations**

As a direct result of Absci leading many of these advancements in AI for biology- we've identified areas of opportunity for policymakers to seize the positive benefits of AI in biology, create fit-for-purpose safeguards, and maintain US leadership in AI and healthcare innovation.

### **1. Lead Today or be Led Tomorrow**

Generative AI might be the fastest-moving field in the history of technology. The United States must lead today, or we will be forced to follow tomorrow. The key to winning this AI leadership race is data.

Congress took a bold and positive step through the CHIPS and Science Act to, in part, invest in the infrastructure needed to win the AI race. And while infrastructure is an important first step- data is the critical next step to truly unlock the potential of AI in many high-impact fields such as biology. Unlike ChatGPT or DALL-E which has the entire internet as training data, biological data is dwarfed in comparison and is the bottleneck to scaling AI in biology. Only by scaling high-quality biological data with platforms such as Absci's can we win the race in AI for biology.

To develop that data infrastructure will require sizable resources, investment, and scalable platforms like Absci to train and run the program's generative AI models. Referencing established natural language scaling laws from MT-NLG/ ChatGPT, we estimate compute spend for training to cost ~\$85M. Additionally, to provide our AI models with the most comprehensive data possible, analogous to the internet for ChatGPT, additional investment in data is required. As an example, to scale data for a virus-focused library of 234M antibody-antigen pairs the cost is estimated to be ~\$284M. Developing such a library of data would enable rapid response to a new pandemic or bioterrorism incident, positioning the US to lead the world in these capabilities.

To that end, a coordinated U.S strategy supporting U.S. leadership in generative AI is necessary. For example, we must invest in creating and securing data to train our nation's AI models, providing U.S. innovators access to resources available to foreign competitors like China.

There is another angle to U.S leadership in AI which I believe is about safeguarding our national health. By pushing the envelope in AI, we're not only accelerating drug discovery and development, but also ensuring we can respond rapidly to health crises and outbreaks.

AI is already beginning to enhance discovery, manufacturing, and clinical trials to ensure a constant pipeline of essential medicines, which minimizes weak spots in our crucial health supply chains. With AI's ability to quickly analyze vast datasets, we can detect adverse drug reactions and potential health threats earlier than ever, protecting our people from unexpected health risks.

In short, being at the forefront of AI isn't just about innovation, it's about building a resilient and adaptive healthcare infrastructure. It's about fortifying our health security against both present and future challenges.

## **2. Safeguarding Against the Weaponization of AI**

As much as AI in biology can be used to transform the lives of patients for the better, without proper safeguards this same technology in the hands of bad actors can be weaponized to do harm at unprecedented levels. Absci's vision is creating breakthrough medicines at a click of a button. By the same principle, bad actors have the potential to leverage AI to create biowarfare and bioterrorism agents at a click of a button. The consequences of that would be devastating for America and American society.

The critical safeguard against the threat of bioterrorism is once again- data. We need a U.S strategy to create and maintain data superiority that positions U.S. innovators to have at their disposal a superior set of AI tools in comparison to terrorist organizations or other opposing forces. The U.S. should implement policies aimed at making data more difficult to access for opposing parties and accelerating the growth of U.S. based biological datasets. The government can play a key role in mitigating harms through collaborations and partnerships with and between U.S innovators. If we look at Operation Warp Speed, American innovation brought the world three lifesaving Covid vaccines in record time. Let's borrow from those winning ingredients to accelerate a U.S Era of AI leadership through focused partnerships between the government and U.S innovators aimed at solving a massive societal issue that affects us all- pandemic preparedness and the protection of U.S citizens from the threat of bioterrorism.

## **3. One-size AI policy Does Not Fit All**

The unique nature of AI in biology vis a vis the general discourse around AI calls for a nuanced approach by policymakers. Some advocate that generative AI should be heavily regulated to prevent its potential misuse. Others think it should be lightly regulated so as not to stifle innovation and public benefit. An all-or-nothing approach like this may be the wrong way to think about it. That's because generative AI introduces different kinds of challenges to different industries.

For example, AI-designed drugs go through the same rigorous FDA approval process as drugs designed by conventional drug discovery campaigns to ensure their safety and

effectiveness, so additional regulations on the AI (which in this case is a tool to discover the drug candidate) would amount to double regulating. This could result in stifling the very innovation that could improve the efficiency and success rates of getting new medicines to patients. On the other hand, it is important that we ensure we have access to high-quality training data that is free from biases that may discriminate against underrepresented groups. Regulators could help establish guidelines to ensure the quality of the underlying data used in designing AI drugs.

## **Summary**

Generative AI might be the fastest-moving field in the history of technology. The United States must lead today or be forced to follow tomorrow. The stakes are too high for the government or industry to tackle AI policymaking alone. For our part, Absci can help guide policy constructively and positively at the intersection of AI and biology. It's also important to me to remind us all not to lose sight of the huge opportunity for patients and society. Every minute of every day, patients wait for a cure, and we have a technology that could potentially deliver breakthrough therapeutics at the click of a button.

It's an immense honor to be able to offer perspective from Absci's work in AI drug creation to the important and time-sensitive dialogue around AI and policymaking. I want to thank Leader Schumer, and Senators Rounds, Heinrich, and Young once again for inviting me. We look forward to continuing to share our insights in AI and biology in helping to guide generative AI toward the greatest public good.

Sean McClain  
Founder & CEO, Absci