Genetic Engineering and Generative AI:

An Explosive Mix

The field of plant biotechnology is undergoing a profound transformation. The rise of generative artificial intelligence (AI) tools is fundamentally reshaping the way plant genetic engineering is conducted. On one hand, AI enhances the precision and efficiency of CRISPR-based gene editing, expanding its capabilities beyond customary gene knockouts. On the other hand, AI-driven genetic engineering may be vulnerable to well-known limitations of AI, such as the black box effect, hallucinations and data errors, raising concerns that plants with undesirable traits could be engineered and released into the environment. This briefing summarises the key findings of Save Our Seeds' report **When Chatbots Breed New Plant Varieties** *and offers our perspective on how the EU should address these emerging issues.*

AI models trained in the 'languages' of biology

Developers have used the AI architectures of the diffusion and large language models used in chatbots like ChatGPT or image generators like DALL-E and trained them in the 'languages' of biology – specifically, protein and genome sequences.

This has become possible because an immense wealth of data on DNA and RNA sequences, proteins and metabolites of plants has become available in recent years. These data now form the raw material that allows the development of generative AI for genetic engineering.

The resulting AI tools are both descriptive and generative. Like conventional deep learning algorithms, they can analyse biological data and make predictions. Additionally, they allow for the design of functional DNA, RNA, and protein sequences, including 'new-to-nature' sequences that have never been observed in nature.

Depending on the type of 'language' used to train the models, several distinct categories have emerged:

- **Protein Models:** Although the use of AI models trained on protein data is a relatively recent development, the number of such tools is growing rapidly. These models can analyse proteins, simulate their interactions and redesign their functions. One of the most notable tools is Google's Alphafold. In recognition of its groundbreaking contribution to the field, Demis Hassabis, head of Google's AI division, along with two colleagues, was awarded the 2024 Nobel Prize in Chemistry for the development of this model.

- **DNA Models:** The first large language models trained on DNA sequences emerged in 2021. Today, four models exist that have been trained specifically on plant DNA. The most advanced of these is AgroNT, a collaboration between Google and Instadeep, released in late 2023. This model was trained on 10 million genome sequences from 48 plant species.

- **RNA Models:** AI models trained on human RNA sequences already exist, and it is expected that plant-based RNA models will be developed in the near future. Models like scGPT, based on single-cell RNA sequencing (scRNA-seq) data, are considered particularly promising for advancing plant science and breeding.

- **Multimodal Models:** Rather than working with a single type of data, AI developers are now working on multimodal models that can process various kinds of biological data. In 2024, Instadeep and BioNTech presented the first multimodal AI architecture capable of connecting DNA, RNA, and protein data.

Use of AI in the genetic engineering of plants

Gene editing today primarily relies on the CRISPR-Cas method. Specific AI tools are now available to enhance the process. These tools assist researchers in identifying optimal targets, suggesting the most effective sequences for the guide RNA and selecting the most suitable CRISPR cutting enzymes. The use of these tools can make gene editing with CRISPR more precise and efficient.

Arguably, AI tools have also expanded the capabilities of CRISPR beyond traditional applications. Developers can now induce not only loss-of-function mutations (knockouts) but also mutations that make it possible to control levels of gene expression. This advancement, known as quantitative trait engineering, is made possible largely thanks to AI. The ability to control gene expression opens up the potential to influence complex quantitative traits, offering new possibilities for plant genetic engineering.

Here are some examples of how AI models can enhance the genetic engineering of plants:

- The US company TreeCo is working to engineer **poplar trees** with reduced lignin production, which could simplify paper production. The company has developed an AI tool that predicts how changes to 21 genes involved in lignin synthesis will impact the trees' wood composition, growth rate, and other traits. The tool identifies over 69,000 potential editing strategies for these genes, narrowing them down to the best options through computer simulations. Based on this analysis, TreeCo is experimentally testing the seven most promising gene-editing combinations in the poplar genome.

- Another US company, Inari, is developing **maize** varieties with reduced height and increased leaf biomass. The company uses an AI tool to predict how mutations in promoter regions will influence plant characteristics. Currently, Inari is conducting field trials of a short-growing maize variety in Belgium.

- Academic researchers have leveraged the Alphafold protein model to redesign patatin, a protein naturally found in **potatoes**. The AI-generated version of patatin is designed to improve the viscosity and nutritional properties of dough made from potato flour. The researchers aim to incorporate this AI-enhanced protein into the potato genome to achieve these improvements.

Large seed companies such as Corteva, Bayer, BASF, and Syngenta are increasingly using AI tools in their genetic engineering programmes. To complement their in-house AI expertise, these companies are also partnering with specialised firms. For instance, BASF and Corteva have initiated collaborations with Tropic Biosciences, which owns proprietary AI technology. Syngenta has teamed up with Instadeep and Biographica, while Bayer is supporting startups Ukko and Amfora, both of which combine AI and CRISPR technologies to develop new plant varieties.

What's next?

The development of generative AI models for gene editing is still in its infancy. Many of the design tools currently available are so new that there is insufficient experimental data to fully assess the performance of their algorithms. However, it is already clear that these tools are creating new design possibilities that go beyond natural limits.

In the coming years, the quality of data acquisition techniques, the volume of data collected and the computing power to process them are expected to grow exponentially. The descriptive and generative capabilities of AI are constantly improving. Experience with large language models trained with microbial DNA sequences demonstrates the potential that genomic AI tools could have. One such model, EVO, can generate sequences on the scale of entire microbial genomes, according to its developers.

As in many other areas, these advances are expected to bring profound changes not only to the life sciences as a whole but also to plant breeding in particular.

What could go wrong?

The integration of AI into genetic engineering raises a number of concerns, many of which reflect the challenges associated with the use of generative AI in other industries. These include, but are not limited to:

- **Lower skill threshold.** Traditionally, plant genetic engineering has been the domain of highly trained professionals. However, with the advent of AI tools, gene editing could soon become accessible to students, computer scientists, entrepreneurs, or even DIY biologists.

- **Black box**. Generative AI models produce predictions or recommendations without providing insight into how or why they arrived at those conclusions. In sensitive areas such as plant genetic engineering, where the products reproduce and interact in nature, and the consequences can affect public health and the environment, the lack of traceability and reproducibility is particularly concerning.

- **Hallucinations.** Generative AI models sometimes produce outputs that seem plausible but are factually incorrect or irrelevant. The frequency and context in which these 'hallucinations' occur, and how to mitigate them, remain unclear.

- **Data distortion.** The outputs and predictions of generative AI models are shaped by the data used to train them. If these data contain errors or biases — whether originating from the biological systems themselves or from human curators — this can be reflected in the model's results.

The lack of specialist skill, in conjunction with the black box, hallucinations and possible data errors, raises concerns that plants with undesirable traits may be engineered and released into the environment. This underscores the need to advance with caution and develop strict oversight.

EU set to deregulate AI-designed plants

At this critical moment, the EU is moving to relax regulatory requirements for the commercialisation of genetically engineered plants. In a proposal from July 2023, the European Commission suggests that plants modified with gene-editing tools like CRISPR should be treated similarly to conventionally bred plants. Specifically, plants with no more than 20 targeted changes to their genome would be exempt from EU GMO regulations. According to the Commission, these plants could be marketed without the need for risk assessments, detection methods, traceability, or mandatory consumer labelling.

Numerous scientists, authorities, and NGOs have criticized the Commission's proposal. The Federal Agency for Nature Conservation (BfN) in Germany highlighted that most gene-edited plants would be released into the environment without any risk assessment, warning that even small genetic changes could have significant consequences and pose high risks. The French food authority ANSES argued that the threshold of 20 nucleotides is not suitable for demonstrating equivalence to conventionally bred plants. In contrast, the European Food Safety Authority (EFSA) defended the Commission's proposal.

Against this backdrop, the convergence of AI and genetic engineering could exacerbate the existing problems with the Commission's proposal. The use of generative AI models may allow developers to fully exploit the 'design space' of 20 genetic modifications, potentially leading to the intentional or unintentional creation of plants that are unsafe for humans and the environment. For instance, researchers could engineer a plant that produces a variety of insect toxins. Under the European Commission's proposal, however, no tests would be required to assess the plant's impact on non-target species, despite the potential for undesirable effects.

The way forward

Instead of relaxing regulatory standards, the EU should uphold the core requirements of its GMO legislation for plants developed using the latest technologies. Mandatory safety assessments should be adapted to account for the unique characteristics of these new processes and technologies.

Furthermore, the EU should take steps to effectively regulate the AI-mediated genetic engineering of plants, as well as other organisms and genetic elements.

Regulations should ensure that the AI models used are reliable and capable of making safe recommendations, while preserving human comprehension, oversight, and decision-making at critical stages of the genetic engineering process.

The containment of AI tools and technologies used in genetic engineering, and of AI-generated artificial organisms, is essential for safe research and development in this field. Monitoring, traceability, and reversibility should be minimum requirements before releasing such organisms into the environment.

International controls should be established to prevent the creation of new organisms or genetic materials that are pathogenic or pose other severe threats. Access to high-risk technologies, tools, and genetic data that could be vulnerable to misuse must be tightly controlled.

Biological safety must be an integral part of all research activities, regardless of whether projects are led by private companies or public research organisations. In cases of uncertainty regarding potential high risks, lower-risk alternatives should be prioritised.

Finally, public funding should be made available to support independent research into the risks of genetic engineering, with particular focus on AI-driven genetic engineering. Special attention should be given to systemic and long-term impacts that go beyond the scope of individual projects (technology assessment).

Save Our Seeds
c/o Foundation on Future Farming
Marienstr. 19-20
10117 Berlin
www.saveourseeds.org/en

Contact:
Franziska Achterberg
Head of Policy and Advocacy
+32 498 362403
achterberg@saveourseeds.org

Design:
Beatriz Francisco
www.linkedin.com/in/beatriz-francisco/

Publication:
January 2025

SØS
save our seeds