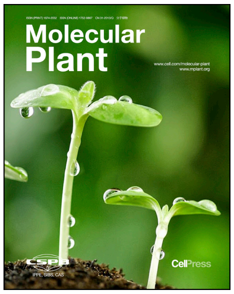


Synergizing Genome Editing and Artificial Intelligence for Predictive Crop Design

Zhaoxu Gao, Jinjie Zhu, Chuanxiao Xie

PII: S1674-2052(26)00116-4

DOI: https://doi.org/10.1016/j.molp.2026.04.007

Reference: MOLP 2121

To appear in: *Molecular Plant*

Received Date: 2 December 2025

Revised Date: 27 February 2026

Accepted Date: 15 April 2026

Please cite this article as: Gao Z., Zhu J., and Xie C. (2026). Synergizing Genome Editing and Artificial Intelligence for Predictive Crop Design. Molecular Plant doi: https://doi.org/10.1016/j.molp.2026.04.007.

Synergizing Genome Editing and Artificial Intelligence for Predictive Crop Design

Zhaoxu Gao [1, #], Jinjie Zhu [1, #], Chuanxiao Xie [1, 2, *]

[1] State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

[2] Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya 572024, China.

Authors/affiliations

Zhaoxu Gao (ORCID: 0000-0002-5398-8054)
State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.
Email: gaozhaoxu@caas.cn

Jinjie Zhu (ORCID: 0000-0002-9844-8043)
State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.
Email: zhujinjie@caas.cn

Chuanxiao Xie (ORCID: 0000-0003-3425-0211)
State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.
Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya 572024, China.
Email: xiechuanxiao@caas.cn

[#] These authors contributed equally to this work.

[*] **Correspondence:** Chuanxiao Xie, State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China. Tel: 010-82107464; Email: xiechuanxiao@caas.cn;

Running title: GE–AI Synergy for Crop Design

Abstract

The convergence of genome editing (GE) and artificial intelligence (AI) is shifting crop improvement from empirical optimization toward predictive design. In this Perspective, we propose that GE and AI are linked by a reciprocal innovation cycle. AI improves GE by enabling more accurate guide RNA design, editing outcome and off-target prediction, and data-driven development of next-generation editing systems. In turn, GE provides a powerful experimental platform to validate AI predictions, dissect causal genotype-to-phenotype relationships, and generate high-resolution perturbation datasets for iterative model refinement. This bidirectional interplay is beginning to accelerate the engineering of complex agronomic traits, including trait stacking, metabolic rewiring, climate resilience, stress tolerance, nutritional enhancement, and *de novo* domestication. Here, we argue that this synergy is emerging most clearly in predictive editing design and iterative model validation, although robust breeding-scale evidence remains limited. We also discuss emerging opportunities for AI-guided closed-loop editing platforms and generative biological design, together with key bottlenecks, including limited training data, biological complexity across scales, model interpretability, and heterogeneous regulatory landscapes. We argue that integrating AI with GE can establish a practical framework for predictive crop design and enable more efficient and sustainable crop engineering.

Short summary

Genome editing and artificial intelligence are jointly transforming crop improvement from empirical breeding toward predictive design. This article highlights how their integration can accelerate trait engineering, optimize breeding decisions, and support the development of resilient, high-performance crops for future agriculture.

Keywords: Genome editing; Artificial intelligence; bidirectional reinforcement; Crop improvement; crop design; foundational framework

1. The GE–AI Co-Evolution in Crop Sciences

Feeding a projected global population of 9 billion by 2050 presents an unprecedented challenge for global agriculture, one further intensified by accelerating climate volatility that threatens to diminish crop productivity and destabilize food systems (Kompas et al., 2024; Lenaerts et al., 2019). Meeting these demands sustainably will require technological breakthroughs that move beyond the inherent limitations of traditional breeding, which remains slow, resource-intensive, and constrained in its ability to predictably engineer complex, multigenic traits (Yuan et al., 2024). The advent of genome editing (GE), particularly CRISPR/Cas systems, has fundamentally transformed plant bioengineering by enabling precise, targeted genetic modifications and markedly accelerating the development of improved cultivars (Chen et al., 2019; Gaj et al., 2013; Gao, 2021). A notable illustration is the editing of promoter regions, which enables precise, quantitative modulation of gene expression without altering coding sequences, thereby improving agronomic traits while minimizing pleiotropic effects. Representative examples include fine-tuning of fruit size and plant architecture without disrupting gene function in tomato (Rodríguez-Leal et al., 2017), which is an empirical study of promoter architecture that demonstrated the principle of quantitative trait modulation through promoter engineering. Regulatory-region engineering of *IPA1* in rice to decouple the trade-off between panicle architecture and tillering (Song et al., 2022). While not AI-driven, these works provide the foundational knowledge that AI models can now leverage to predict optimal promoter targets. Unlike traditional coding-region editing, which aims to obtain qualitative traits through loss or alteration of function, promoter editing achieves fine-tuning gene expression levels by precisely regulating cis-regulatory elements (CREs). This approach generates continuous, predictable quantitative trait variations and typically does not completely disrupt gene function.

In parallel, AI has become central to advancing GE. Machine learning (ML) approaches now underpin key steps in the GE workflow, including optimizing guide RNA (gRNA) design, predicting editing efficiencies, and reducing off-target risks (Chuai et al., 2018; Huang et al., 2023; Wang et al., 2019). Beyond tool enhancement, AI's capacity to integrate and interpret large-scale multi-omics datasets is poised to substantially strengthen breeding pipelines and contribute to global food-security solutions (Farooq et al., 2024). Importantly, the convergence of AI and GE establishes a dynamic feedback system: AI-driven predictions inform editing strategies, while GE-derived outcomes generate high-fidelity empirical data that validate, constrain, and continually improve computational models (Dixit et al., 2024; Sulis et al., 2023; Zhang et al., 2023b; Zhao et al., 2025).

This convergence of molecular precision and computational prediction represents a transformative synergy for crop improvement (Figure 1). In this Perspective, we use “AI–GE synergy” in a narrower sense to refer to cases in which AI models contribute to the prediction, design, or optimization of editing outcomes, and genome editing in turn provides experimental data for iterative model improvement. Yet its full potential remains tempered by persistent challenges, including model brittleness (Table 1**:** the tendency of AI models to fail when presented with data that deviates from their training distribution), gaps in high-quality training data, and regulatory fragmentation across jurisdictions. In this perspective, we trace the development of this emerging integration, examining the bidirectional relationship between AI-driven acceleration of genome editing technologies and the use of GE as a biological validation framework for AI models (Figure 1). We first outline AI’s role as a computational engine for editor design, optimization, and performance forecasting, and then examine how GE generates empirical datasets that ground-truth and iteratively refine these models. Through selected case studies, along with an assessment of emerging opportunities and remaining obstacles, we identify future priorities for advancing equitable, scalable, and sustainable agricultural innovation. Key concepts underlying the GE-AI synergy are summarized based on our assessment of current field practices in Table 1 (Glossary of Key Terms). Importantly, not all computational or high-throughput approaches associated with genome editing should be considered evidence of true GE-AI synergy. In this Perspective, we use the term more specifically to describe situations in which AI contributes to the prediction, design, or optimization of editing outcomes, and genome editing in turn generates experimental data that can iteratively refine these models. We therefore distinguish genuine AI-enabled editing frameworks from broader digital tools, rule-based design systems, or screening platforms that support genome editing workflows but do not yet constitute closed-loop AI-guided crop design. Under this definition, the strongest current evidence lies in predictive editing design and model-guided optimization, whereas many breeding-oriented applications remain emerging rather than fully established.

Table 1. Glossary of Key Terms in GE–AI Synergy for Predictive Crop Design.

Term	Definition
AI Concepts	
Artificial Intelligence (AI)	Computational methods that learn from data to predict, classify, or design biological outcomes.
Machine Learning (ML)	A subset of AI that identifies patterns from data to make predictions or decisions without explicit rule-based programming.
Deep Learning	A branch of ML that uses multi-layer neural networks to learn complex patterns from large datasets.
Explainable AI (XAI)	Methods that help interpret how AI models generate predictions or decisions.
Federated Learning	A distributed learning approach in which models are trained across multiple sites without sharing raw data.
Hallucination	AI generation of plausible-looking but biologically invalid or non-functional outputs.
Model Brittleness	Reduced model reliability when applied to data that differ from the training distribution.
Model Architectures	
CNN / Transformer / Hybrid	Neural network designs tailored to biological sequence features. **CNNs** excel at detecting local motifs (e.g., PAM-proximal nucleotides for sgRNA efficiency); **Transformers** capture long-range dependencies (e.g., enhancer-promoter loops); **Hybrid models** (e.g., StripedHyena) balance local precision with global context.
Generative Design	
Diffusion Models	Generative models that create new structures by learning to reverse a noise-adding process.
De Novo Design	Computational creation of novel biological sequences or molecules from desired functional specifications rather than natural templates.
Genome Editing and Crop Design	
Genome Editing (GE)	Technologies enabling precise DNA modifications using programmable nucleases; the experimental execution layer of the GE-AI synergy.
CRISPR/Cas Systems	RNA-guided editing toolkit (Cas effector proteins + guide RNAs); AI-optimizable at both effector and targeting components.
Off-Target Effects	Unintended genomic modifications at non-target sites; AI mitigation via learning mismatch tolerance patterns and chromatin accessibility.
GE–AI Synergy (Closed-Loop)	An iterative framework in which AI-guided design is tested by genome editing and the resulting data are used to improve subsequent models.

Multi-Omics Data	Integrated molecular datasets (genome, epigenome, transcriptome, metabolome) that serve both as training inputs for AI and as validation metrics for GE outcomes.
Trait Stacking	The simultaneous introduction or optimization of multiple agronomic traits. AI optimizes genetic combinations to avoid negative epistasis and balance pleiotropic trade-offs.
De Novo Domestication	Rapid conversion of wild species to elite crops via editing domestication genes; AI-enabled target identification and multiplex strategy design.

2. AI-Assisted Computational Engine for GE

Artificial intelligence (AI), particularly its subset of machine learning (ML), is transforming genome editing by providing the computational power necessary to extract meaningful patterns from increasingly large and complex omics datasets (van Dijk et al., 2021). Modeled loosely on biological neural architectures, modern AI systems excel in predictive modeling, functional inference, and *de novo* biological design, thereby offering capabilities that markedly enhance the precision, efficiency, and scalability of genetic engineering workflows (Richards et al., 2022; Richards et al., 2019). In this section, we examine how AI functions as a computational engine for GE along two primary axes: optimizing targeting strategies to improve editing accuracy and mitigate risk, and enabling the rational design and refinement of genome-editing machinery itself.

2.1. AI Model Architecture Comparison and Biological Utility

The rapid development of AI has given rise to a diverse ecosystem of model architectures applied to genome engineering and functional genomics. As a powerful automatic feature extractor, CNN (see Table 1) excels at recognizing local dependency patterns and spatial hierarchical features from one-dimensional sequence data (Alipanahi et al., 2015). This makes it highly advantageous for automatically learning key sequence motifs, capturing position-specific effects (e.g., the influence of nucleotides proximal to the PAM site), and performing end-to-end efficiency prediction (e.g., sgRNA sequences) (Chuai *et al.*, 2018). Consequently, CNN became the mainstream underlying architecture in the early development of gene editing prediction tools.

Transformer, equipped with self-attention mechanisms, employs capturing long-range dependencies between any positions within a sequence (Avsec et al., 2021; Rives et al., 2021). This directly addresses the complex biological characteristics of "long-distance site interactions" in biological sequences, such as enhancer-promoter interactions on the genome or tertiary structure interactions in proteins. This renders it well-suited for the prediction of off-target effects, the design of highly specific gRNAs, and the annotation of functional elements within large crop genomes. Nevertheless, these models typically require substantial computational resources and large-scale training datasets—resources that remain in short supply for non-model crop species.

While CNNs and Transformers offer complementary strengths, integrating both approaches enables hybrid architectures that balance local precision with global context-awareness. For instance, StripedHyena, used in the Evo model, combines the efficiency of convolutions with the context-awareness of attention mechanisms to tackle multi-scale biological problems (Nguyen *et al.*, 2024). In addition, protein language models (e.g., ESMFold, ProGen) leverage evolutionary-scale sequence data to capture latent structural constraints, offering an alternative paradigm for *de novo* protein engineering (see Section 2.2). These models preserve interpretability while sustaining high predictive performance, which is a critical equilibrium for applications in gene editing, as reliability and transparency directly impact experimental validation.

This architectural diversity empowers researchers to match tools to specific biological tasks: CNNs remain sufficient for high-throughput sgRNA screening, where local PAM-proximal features dominate; Transformers are preferable for promoter engineering requiring long-range chromatin loop modeling; Hybrid models and language models excel in *de novo* enzyme design demanding multi-scale constraints or evolutionary context. The choice of architecture ultimately depends on the specific task, the volume of available data, and the requirements for interpretability.

A coherent evaluation framework is therefore essential to fairly compare architectural strengths and weaknesses. Key benchmarks should include not only classical predictive metrics such as precision, recall, F1-score, and AUC, but also task-specific criteria including gRNA efficiency correlation with wet-lab validation, off-target prediction accuracy, inference speed on gigabase-scale genomes, and generalization performance across cultivars, tissues, and experimental conditions. Without standardized benchmarks, even sophisticated architectures risk being validated in limited settings, limiting their translational value to crop breeding and genome engineering.

2.2. Optimization and *De Novo* Editor Design

2.2.1. *De Novo* Design and Discovery of Novel Editing Systems

A major frontier in AI-driven genome engineering is the refinement and *de novo* design (see Table 1) of genome editors, which shifts the role of AI from prediction to generative creation (Figure 1). The breakthrough in protein structure prediction, exemplified by AlphaFold, has yielded highly accurate structural models for vast numbers of proteins, including those central to genome editing systems (Abramson et al., 2024; Tunyasuvunakool et al., 2021). RoseTTAFold provides robust, open-source protein structure prediction to support the engineering and refinement of Cas nucleases and accessory proteins (Humphreys et al., 2021), while ESMFold delivers high-speed, accurate folding predictions for single sequences, making it particularly useful for iterative design workflows (Lin et al., 2023). Building on these advances, diffusion models (see Table 1) such as RFdiffusion now operate as 'structure-generation' engines capable of end-to-end *de novo* design of protein scaffolds and functional domains, thereby enabling the creation of entirely new classes of genome-editing enzymes (Watson et al., 2023).

These structural insights have also enabled AI-driven mining of natural diversity. Tools such as FLSHclust and CHOOSER have uncovered previously unrecognized CRISPR systems, including self-processing Cas12 variants and Casλ homologs, thereby extending the available genome-editing toolkit (Altae-Tran et al., 2023; Li et al., 2024). These successes underscore AI’s strength in integrating structural and sequence-based information to detect distant homologies and functional relationships that traditional alignment approaches miss. AI-driven exploration of protein space is continually revealing new systems and functional components. For example, structural modeling has provided key insights into the evolutionary origins of Cas13 and has identified novel proteins suitable for next-generation base-editing systems (Huang *et al.*, 2023; Yoon et al., 2024). Notably, AI-guided mining recently identified the TranC family of RNA-guided nucleases, illuminating the evolutionary transition from transposons to Cas12-like systems and revealing how RNA modularity drives CRISPR architectural innovation (Jin et al., 2025).

Moving beyond natural evolution, deep learning models such as ProGen and EVOLVEpro now enable *de novo* design and creation of proteins with tailored biochemical properties, unlocking forms of genome manipulation that were previously unreachable by natural evolution alone (Jiang et al., 2024; Madani et al., 2023). This capability has already produced new CRISPR systems—including EvoCas9-1 and OpenCRISPR-1—that match or exceed the performance of their naturally occurring counterparts (Nguyen *et al.*, 2024; Ruffolo et al., 2025). By coupling structural prediction with generative modeling, AI is now enabling the creation of highly specific, multifunctional genome editors designed to meet diverse and complex engineering needs.

2.2.2. Optimization and Refinement of Editing Tools

The same AI-driven approaches extend to the enhancement of existing editing machineries. For instance, ProMEP (Huang *et al.*, 2023) was used to engineer the TnpB protein, increasing its editing efficiency threefold, while AI-guided optimization of Cre recombinase achieved a 3.5-fold boost in recombination activity (Fei et al., 2025). Specifically, the AiCE (AI-formed Constraints for Protein Engineering) framework, an approach that integrates structural and evolutionary constraints, was used to model the evolutionary trajectories of several genome-editing enzymes (Fei *et al.*, 2025). AI-driven approaches have also contributed to innovations such as traceless Re-pegRNA architectures, which enable scarless, megabase-scale genome manipulation (Sun et al., 2025). Collectively, these developments highlight how AI can refine and expand genome-editing systems, improving precision and efficiency while lowering developmental barriers and cost.

Complementing these advances, an expanding set of AI-enabled tools is broadening the capabilities available to genome engineers. Notably, specialized models can now generate custom-designed zinc-finger arrays, providing a programmable alternative to CRISPR-based systems and further diversifying the molecular toolkit for targeted DNA recognition (Ichikawa et al., 2023). The integration of these computational platforms is catalyzing a new phase of innovation, improving both the design and functional performance of genome-editing reagents.

2.3. Outcome Prediction, Precision Targeting, and Risk Mitigation

AI optimization of genome editing extends from enzyme engineering to the precision deployment of these tools *in planta*. Three computational capabilities are of paramount importance: predicting editing outcomes to maximize efficiency, designing optimal single-guide RNAs (sgRNAs) for precise targeting, and mitigating off-target effects to ensure specificity. Together, these functions bridge the gap between the design of molecular tools and the achievement of desired phenotypes in crop systems.

2.3.1. AI-Guided Design of sgRNAs and Editing Outcome Prediction

A more profound shift is underway from predicting the efficacy of conventional designs to the *de novo* generation of optimized sgRNAs. Tools such as sgRNAGen employ deep generative models to create sgRNAs *de novo*, producing highly efficient sequences without relying on natural templates (Xia et al., 2025). Large language model-based methods advance this capability through deep learning of sequence–function relationships, enabling precise prediction of functional sgRNAs for various Cas9 variants while significantly improving both efficiency and specificity (Ruffolo *et al.*, 2025). This shift from rule-based selection to *de novo* guide RNA design represents a technological leap in genome editing, substituting trial-and-error screening with predictive, data-driven workflows.

Besides sgRNA optimization, AI-directed approaches can also guide the engineering of the editing machinery. DeepCas9variants identifies the most effective Cas9 variant for a given genomic site, whereas the PAMmla framework links amino acid composition to PAM specificity, thereby facilitating the development of Cas9 variants with customized targeting profiles (Silverstein et al., 2025). These approaches are particularly valuable in polyploid crops with complex genomes, such as wheat and rapeseed, where targeting flexibility is essential for accessing diverse genomic loci.

AI-assisted prediction of editing efficiencies has evolved from simple statistical models to sophisticated deep learning frameworks. For base editing, DeepBE forecasts outcomes across diverse genomic contexts, enabling researchers to select optimal sgRNAs before experimental validation (Kim et al., 2023). As for prime editing, a modality particularly relevant for precise crop improvement, the ePRIDICT model integrates chromatin accessibility features to predict editing success, addressing the challenge that chromatin states in plants vary dramatically across tissues and developmental stages (Mathis et al., 2024).

2.3.2. Crop-oriented sgRNA Design, Editing-outcome Prediction, and Risk Mitigation

The essence of gene editing lies in precise sgRNA design, which dictates editing efficiency and specificity. While fundamental principles are similar across kingdoms, plant-specific differences in PAM sequences, GC content, and target selection necessitate tailored tools (Lei et al., 2014). More importantly, however, current plant-focused platforms do not solve a single unified prediction problem; rather, they address different layers of the editing workflow, including guide selection, editing-outcome estimation, and off-target risk assessment. This fragmentation partly explains why the field still relies on combining multiple tools rather than on an integrated predictive framework. Crop gene editing prediction has evolved through three phases: statistical modeling, deep learning adoption, and AI integration.

CRISPR-P, an early plant-specific tool trained on Arabidopsis and rice data, remains widely used in model plants (Lei *et al.*, 2014; Liu et al., 2017). CRISPR-GE offers a more comprehensive platform, integrating sgRNA design, primer design, off-target analysis, and mutation detection via Gradient-Boosting Regression incorporating GC content, dinucleotide frequency, and chromatin accessibility (Xie et al., 2017). In parallel, broadly adopted platforms such as CRISPOR and CHOPCHOP remained useful in plant studies because of their accessibility, transparency, and multi-model comparison functions, even though they were not developed specifically for crop systems (Concordet and Haeussler, 2018; Labun et al., 2019)—CHOPCHOP for its user-friendly interface (Labun *et al.*, 2019) and CRISPOR for algorithmic transparency and multi-model evaluation (Concordet and Haeussler, 2018). Taken together, these tools illustrate the first major stage of the field: sequence-centered guide design. Their practical value is clear, but their predictive scope is often constrained by training data derived from a limited number of species or experimental contexts, which reduces transferability across diverse crop genomes.

CRISPR-BETS represents a step toward outcome-aware design for base editors by prioritizing guides predicted to introduce stop codons. Nevertheless, the predictive scope of CRISPR - BETS is restricted by rule - based assumptions and the limited training data. (Wu et al., 2021). Tools like CCTop and Cas-Designer enable sensitive off-target scanning with customizable mismatch and PAM rules, but they primarily prioritize sequence-level specificity rather than in planta editing outcomes (Maas et al., 2015; Park et al., 2015). This reveals a second limitation in current crop editing pipelines: efficiency, product spectrum, and off-target liability are usually modeled separately, even though breeders ultimately need to optimize them jointly in a species- and locus-dependent manner.

The optimal strategy involves cross-utilizing multiple tools: initial screening with CRISPR-P 2.0 or CRISPR-GE, efficiency evaluation via CRISPOR or CHOPCHOP, and off-target validation with CCTop. A central challenge is that most available models remain sequence-aware but not sufficiently context-aware: they rarely incorporate chromatin state, repair pathway bias, transformation system effects, genotype dependence, or species-specific editing behavior at a scale relevant to crop improvement. Next-generation AI-driven platforms will transform this landscape by using deep learning to predict sgRNA efficiency in crop cells, integrating rule-based outcome prediction, and learning composite specificity scores, including off-target effects. More importantly, future systems will need to move beyond isolated prediction tasks toward unified models that jointly estimate editability, expected editing outcomes, and biosafety risk in realistic crop backgrounds.

To conduct a meaningful evaluation of these plant-specific AI tools, a structured benchmarking framework needs to be established. This framework should delineate standardized datasets that cover both model and non-model crop species, consistent performance metrics, and evaluation criteria that account for cross - species transferability, computational efficiency, and usability within practical breeding pipelines. Only via systematic and reproducible validation can plant - specific AI tools evolve from specialized software components into reliable and widely adopted instruments for precision genome editing in agricultural applications. The long-term goal is therefore not merely better sgRNA design software, but a context-aware and benchmarked predictive framework for robust genome editing across diverse crop systems.

2.3.3. Comprehensive Prediction and Mitigation of Off-Target Effects

Off-target activity (see Table 1) remains a major barrier to the reliable application of genome editing technologies in crops, particularly in polyploid species where multiple homoeologous loci must be considered simultaneously. AI now enables genome-wide assessment of both on-target and off-target effects across CRISPR/Cas systems, base editors, and prime editors. Tools such as CRISOT provide a unified framework for evaluating editing specificity by integrating RNA–DNA interaction fingerprints to predict unintended cleavage sites (Chen et al., 2023). In plant applications, these predictive capabilities are being extended to address species-specific challenges. For example, AI-driven screening of transposon elements has identified candidates suitable for hybrid systems that enable highly accurate, RNA-delivered insertion of large DNA fragments (Fei *et al.*, 2025). Together, these advances strengthen the precision and translational potential of genome editing by ensuring that engineered modifications are both efficient and specific.

By integrating these capacities of design, prediction, and risk mitigation, AI is transforming genome editing from a trial-and-error discipline into a predictive engineering workflow. This integration lays the groundwork for the next generation of precision editing platforms, where models trained on diverse genomic and epigenomic contexts will enable the reliable and specific modification of even the most complex crop genomes.

3. GE as the Biological Validation Platform for AI-assisted Predictions

Despite remarkable advances, the iterative improvement of AI models is fundamentally constrained by the scarcity of high-quality biological validation data. GE could ideally function as the definitive testing ground to overcome this limitation, translating *in silico* predictions into biological reality. Each edited organism becomes an empirical benchmark, yielding quantitative fitness data, high-resolution mutation readouts, and multi-omics profiles that iteratively refine, constrain, and stress-test AI models against the inherent complexity of living systems.

3.1. Bridging *In Silico* and *In Planta*

3.1.1. Validating Multi-Gene Editing Strategies Across Plant Systems

The progression from computational prediction to measurable organismal trait offers one of the clearest demonstrations of GE-AI synergy. A striking example comes from forestry, where machine learning was used to evaluate nearly 70,000 candidate editing strategies across 21 monolignol biosynthesis genes. From this expansive *in silico* screen, seven optimal multi-gene strategies were prioritized for empirical testing. The resulting CRISPR-edited trees exhibited a 35% reduction in lignin content and more than a 200% increase in both carbohydrate-to-lignin (C/L) and syringyl-to-guaiacyl (S/G) ratios, thereby validating the model's predictions and offering a blueprint for sustainable fiber production (Sulis *et al.*, 2023).

Beyond forestry, multi-gene-editing frameworks have been extended to major crop species. For example, the development of the BREEDIT platform in maize enables rapid generation of combination mutants through a high-throughput gene-mining pipeline, thereby accelerating the identification of candidate genes and gene combinations for complex trait improvement (Lorenzo et al., 2023). Similarly, whole-genome sequence analysis in tomato enabled the design of sgRNAs capable of simultaneously targeting multiple homologous genes using the CRISPys algorithm, facilitating the functional dissection of loci underlying fruit shape, size, flavor, and disease resistance (Berman et al., 2025).

3.1.2. Fine-Tuning Gene Expression for Trait Improvement

One notable example study is the application of deep learning models to identify cis-regulatory elements (CREs) in crop genomes, providing a mechanistic foundation for targeted manipulation of gene expression. By incorporating the concept of "editing plasticity", this work enabled the prediction and experimental validation of expression changes in maize, ultimately increasing α-tocopherol accumulation (Qiu et al., 2025). This hybrid approach combining AI and GE demonstrates a scalable path for quantitative trait optimization beyond natural allelic variation.

3.1.3. Engineering Plant Immunity via AI-Discovered Components

AI is also being used to strengthen plant immunity. For example, AI-based screening approaches have identified anti-protein-degradation peptides (APPs) capable of stabilizing the MYC2 transcription factor, a central regulator of immune responses. Experimental validation of these AI-designed peptides demonstrated enhanced tolerance to citrus Huanglongbing (HLB), highlighting the efficiency of AI in uncovering actionable molecular strategies (Zhao *et al.*, 2025). In a related effort, AI-guided engineering of the selective cold shock protein receptor (SCORE) in citrus showed that its recognition spectrum could be broadened to detect cold shock proteins (CSPs) from additional pathogens, including those responsible for HLB (Ngou et al., 2025). While these studies currently rely on peptide synthesis and transient assays for validation, the integration of GE to introduce these AI-optimized modifications as stable, heritable edits represents the critical next step for translating computational discoveries into crop improvement.

3.2. Data Generation for AI Refinement

3.2.1. Multi-Omics Data Fuels Iterative Model Improvement

Beyond validating individual predictions, the datasets generated through GE experiments are essential for the continual improvement of AI models. High-quality multi-omics profiles spanning genomic, transcriptomic, epigenomic, and metabolomic layers from edited plants provide the ground-truth information needed to retrain and refine predictive algorithms. For example, one study integrated diverse multi-omics datasets with AI-based analysis to identify gene combinations associated with drought resistance in maize. Subsequent CRISPR/Cas9 editing confirmed the functional relevance of these targets (Xie and Zhang, 2025). This establishes a virtuous cycle in which each round of genome editing not only validates AI predictions but also enriches the underlying training data, yielding progressively more accurate models (Figure 1).

3.2.2. GE as a Functional Testbed for AI-Predicted Structures

This iterative feedback model is broadly applicable. In another study, AI-identified genes associated with high yield and disease resistance in maize were edited, producing lines with markedly improved agronomic traits (Chen et al., 2025). Similarly, GE provides functional validation for *de novo* protein designs generated by AlphaFold and related platforms, offering direct empirical readouts of editing efficiency that validate structural predictions (Fei *et al.*, 2025; Huang *et al.*, 2023). Collectively, these examples illustrate how GE transforms AI from a strictly predictive framework into a dynamic learning system that evolves through continuous empirical feedback.

3.3. A Validation Framework for GE-AI Integration

A major barrier to meaningful AI–GE integration in crops is the lack of a shared validation framework. Current models are often assessed using heterogeneous datasets, species-specific assumptions, and tool-specific criteria, making it difficult to compare performance or judge biological relevance across studies. To address this gap, we propose a three-tier framework that integrates computational benchmarking, experimental validation, and iterative model refinement.

At the first tier, *in silico* benchmarking should evaluate models for sgRNA activity, off-target propensity, editing outcomes, and regulatory sequence design using standardized crop datasets and harmonized metrics. Beyond conventional measures such as accuracy, precision, and recall, benchmarking should also assess robustness across species, genomic contexts, and training–test splits, thereby revealing the extent to which models are transferable beyond their original datasets.

At the second tier, *in planta* validation should test whether predicted edits retain utility in real crop systems. This step should measure not only editing efficiency and specificity, but also molecular and phenotypic consequences in representative crop genotypes. Cross-species validation in crops such as rice, wheat, maize, and tomato is particularly important for defining the practical boundaries of model generalization.

At the third tier, iterative feedback should use experimental results to refine predictive models. Multi-omics data from edited plants can capture downstream regulatory and metabolic consequences that are invisible to sequence-only models, thereby enabling more context-aware prediction. Repeated cycles of prediction, editing, phenotyping, and retraining could progressively improve model reliability and biological realism.

Under this framework, tools such as CRISPR-P, CRISPR-GE, and DeepBE could be benchmarked using common crop-specific standards rather than isolated case studies. More broadly, such a framework would shift the field from fragmented tool evaluation toward a reproducible and crop-aware foundation for AI-guided genome editing.

4. Emerging Spectrum of GE–AI Synergy in Action: Case Studies

The convergence of GE and AI is moving from a theoretical promise to a demonstrable impact across agricultural systems. We here illustrate a paradigm shift in which AI identifies optimal genetic configurations, GE implements precise edits at scale, and iterative data feedback continually sharpens predictive accuracy (Figure 2).

4.1. Climate Resilience and Nutritional Enhancement

AI-guided GE is delivering rapid solutions for abiotic stress adaptation and nutritional fortification through precise manipulation of individual regulatory nodes. In rice and tomato, prime editing-mediated insertion of heat-shock elements (HSEs) into cell wall invertase (*CWIN*) promoters identified through expression quantitative trait locus (eQTL) modeling enhanced heat-responsive carbon allocation, mitigating yield losses by up to 41% under thermal stress (Lou et al., 2025).

Concurrently, AI-enabled discovery of non-coding regulatory variants is overcoming the constraints of natural genetic diversity. AI-guided interrogation of CoQ biosynthetic pathways identified *Coq1* as the determinant of side-chain length; editing this locus in rice and wheat shifted production from CoQ9 to CoQ10, enriching antioxidant capacity (Xu et al., 2025). The created allele is novel and absent from germplasm banks. Similarly, researchers used the *Basenji2* model to predict that editing a specific 5' UTR variant in *ZmVTE4* would enhance vitamin E accumulation. GE-mediated validation confirmed a substantial increase in kernel α-tocopherol (Qiu *et al.*, 2025). These cases exemplify how GE-AI synergy transcends natural allelic limitations to achieve predictable metabolic engineering.

4.2. Biotic Stress Resistance

Biotic stress, arising from interactions with living organisms such as fungi, bacteria, viruses, nematodes, insects, and parasitic plants, represents a major constraint on plant health, productivity, and ecosystem stability.

A growing body of work demonstrates that genome editing can be used to identify and deploy rare resistance alleles that balance disease resistance with agronomic performance-a long-standing challenge in plant breeding. For example, large-scale screening of genome-edited wheat populations identified the *Tamlo-R32* allele, conferring complete powdery mildew resistance without yield penalties (Li et al., 2022b). AI-based analysis of the *OSRBL1* promoter subsequently predicted optimal cis-regulatory modifications, enabling precise editing that generated a gain-of-function allele (*OSRBL1Δ12*) with broad-spectrum resistance to blast, bacterial blight, and false smut while maintaining yield stability (Sha et al., 2023). These successes underscore AI's capacity to navigate complex genotype-phenotype landscapes that confound conventional breeding, particularly in balancing disease resistance with agronomic performance.

Crucially, the convergence of AI and genome editing opens the door to genuinely synergistic workflows in which predictive models inform precise editing, followed by rapid experimental validation. As such, as closed-loop GE-AI frameworks mature, they are likely to underpin the next generation of crop improvement strategies aimed at durable, adaptive multi-pathogen resistance.

4.3. From Trait Stacking to *De Novo* Domestication: AI as the Navigator of Complex Genetic Landscapes

The most sophisticated expression of GE-AI integration is the coordinated improvement of multiple traits, commonly referred to as trait stacking (see Table 1), and the accelerated *de novo* domestication (see Table 1) of wild species (Figure 2). AI, and particularly reinforcement learning frameworks, is increasingly being used to design multiplex editing strategies that balance yield, disease resistance, nutritional quality, and other agronomic traits that have historically been difficult to optimize simultaneously (Fernie and Yan, 2019; Zhang et al., 2023a).

Proof-of-concept studies have demonstrated rapid domestication through editing of known domestication genes, e.g., *CLV3/WUS* in tomato for fruit size (Zsögön et al., 2018) or the shattering gene in wild rice (Yu et al., 2021), achieving rapid domestication in years that took millennia of conventional breeding. These demonstrated successes rely on rational engineering of well-characterized, monogenic traits. However, the genetic architectures of modern elite crops emerge from polygenic networks involving epistatic interactions, pleiotropic trade-offs, and genotype-by-environment dynamics that transcend rational design. While domestication history is unknown and adaptive traits are dispersed across diverse germplasms, both trait stacking and *de novo* domestication of wild relatives require navigating combinatorial explosions beyond human calculation.

AI-driven crop genetic improvements necessitate three data pillars: (i) pan-genomic atlases to identify cryptic beneficial alleles absent from reference genomes; (ii) phenomic-resolution trait maps linking molecular variation to field performance; and (iii) regulatory network models predicting non-additive epistasis. Armed with Integrative network data, machine learning can optimize multiplex editing strategies, balancing yield, resilience, and nutritional traits, and forecasting G×E interactions in the future.

5. Emerging Frontiers and Technical Hurdles

5.1. Toward Self-Driving Plant Improvement

The convergence of AI and GE is opening the door to a new generation of transformative agricultural technologies (Figure 3). Among the most forward-looking developments are autonomous editing platforms that integrate AI-driven design with robotic execution to automate the entire pipeline, covering target selection, sgRNA design, vector construction, and delivery to phenotyping and data capture. However, advances remain asymmetric across this pipeline. Numerous models have been developed to identify key genes within metabolic pathways (Pei et al., 2025; Wang et al., 2025a; Zhang et al., 2025b), predict optimal protein structures and functions (Wang et al., 2025b; Yao et al., 2025), and infer transcriptional regulatory landscapes in plants, exemplified by tools such as PhytoExpr, Basenji2, and GenoRetriever (Gao et al., 2025; Qiu *et al.*, 2025; Li et al., 2024). In parallel, a growing ecosystem of online platforms supports sgRNA design, prediction of editing outcomes, efficiency estimation, and off-target analysis across diverse genome-editing tools and plant species (Labun *et al.*, 2019; Li et al., 2025a; Lin et al., 2021; Liu *et al.*, 2017), substantially lowering the barrier to in silico design. In contrast, wet-lab steps, including vector assembly and genetic transformation, are still largely manual-operation-driven and lack scalable automation, underscoring the need for coordinated development of intelligent robotics and adaptive experimental agents. Encouragingly, progress in AI-enabled phenotyping, such as three-dimensional deep learning-based analysis in tomato, demonstrates that downstream automation is becoming feasible (Jeon et al., 2025).

Notably, the recent development of CRISPR-GPT demonstrates the feasibility of an end-to-end AI platform that integrates design, experimental guidance, and data analysis, successfully enabling even novice users to perform efficient and precise gene knockout and activation experiments (Qu et al., 2025). These advances suggest that a plant-oriented counterpart, such as a fully integrated Plant CRISPR-GPT platform, may emerge in the near future, bringing true end-to-end automation to plant genome editing. One recently developed high-throughput system, for example, can generate thousands of edited samples within a single week. The extensive, real-time datasets produced by such platforms have already enabled the training of specialized predictive models, including the Chromatin Accessibility Enabled Learning Model (CAELM), which forecasts cytosine base-editing outcomes by incorporating both sequence context and chromatin accessibility (Li et al., 2022a). Looking ahead, these autonomous frameworks point toward the emergence of self-optimizing, closed-loop “design–build–test–learn” workflows capable of continuously improving their own predictions and output (Figure 1). Such systems lay the groundwork for the development of SMART crops—plants enhanced with sensing, monitoring, and responding genetic circuits that dynamically adjust their physiology to maximize yield or stress resilience. The integration of AI-guided synthetic biology will be central to enabling these adaptive, high-performance crop systems (Zhang et al., 2025a).

5.2. Generative AI and the Hallucination Challenge

Beyond natural variations, generative AI, particularly diffusion models and protein language models, now enables the *de novo* design of genome editors and synthetic regulatory circuits. Generative AI is expanding synthetic biology toolkits, with large language models like GPT-4 extracting structured knowledge to generate datasets for predicting complex traits such as microbial fermentation yield (Xiao et al., 2023). However, achieving dynamic, functional protein designs requires next-generation tools incorporating biophysical constraints, including folding pathways and conformational dynamics (Winnifrith et al., 2024). These approaches can explore sequence contexts unreachable by natural evolution, creating novel enzymes with tailored PAM specificities or catalytic activities (Li et al., 2025b; Notin et al., 2024)—yet challenges persist in multi-component assembly design, difficult epitope targeting, and predicting protein-metabolite interactions

(Wang et al., 2022; Wu et al., 2025; Zambaldi et al., 2024). Future advances hinge on integrating sophisticated biophysical constraints to enable robust in vivo protein and regulatory circuit design.

However, this capability introduces the critical challenge of biological hallucination (see Table 1): structures that appear energetically stable *in silico* yet fail to fold or function *in vivo* (Jumper *et al.*, 2021). To mitigate such risks, a multi-tiered quality control strategy must be established throughout the design process. Tools like ProteinMPNN (Dauparas et al., 2022) can be used to design more stable and more readily expressed amino acid sequences for the generated structures. Functional computational validation, including flexible docking and free-energy calculations, assesses the potential for interactions with target molecules (Azoitei et al., 2011). Ultimately, high-throughput wet-lab validation, comprising expression and purification, biophysical characterization, and functional assays, serves as the gold standard, forming a rational, closed-loop screening pipeline from computation to experimentation (Wicky et al., 2022).

Looking ahead, developing hybrid models that incorporate physical constraints, constructing enhanced datasets that include dynamic functional information and negative feedback, and establishing an automated "design-build-test-learn" experimental loop will be key to advancing generative AI from "pattern imitation" toward "reliable biological creation".

5.3. Protein-coding Region Prediction Challenges

AI applications in plant genome editing are expanding from target identification toward optimizing the entire design chain, with promoter editing emerging as a particularly promising direction. Unlike complete gene disruption, promoter editing offers higher safety and flexibility (Rodríguez-Leal *et al.*, 2017). Deep learning models integrating chromatin accessibility and epigenomic data can efficiently predict regulatory regions, enabling optimal target identification (Avsec *et al.*, 2021).

However, predicting phenotypic outcomes from coding mutations—especially non-synonymous ones—remains exceptionally complex due to multiple factors: altered codon usage and translation efficiency (Buhr et al., 2016), impacts on protein structure and stability (Budisa et al., 2023; Jumper *et al.*, 2021), and context-dependent epistatic interactions between mutations (Starr and Thornton, 2016). These layers collectively create significant uncertainty in the prediction of protein-coding region editing to phenotype.

Current editing strategies follow two complementary paths: "knowledge-driven" approaches leverage existing structural, functional, or evolutionary insights—such as applying validated animal homolog information to plant targets (Li et al., 2018)—while "AI-driven" methods use machine learning to uncover patterns from genomic and phenotypic data. The most effective strategy combines both: employing domain knowledge as constraints to guide AI exploration within biologically plausible spaces, thereby accelerating the Design-Build-Test-Learn cycle and enabling precise, predictable genome editing (Figure 1). This represents a paradigm shift from "disruptive" toward "regulatory" editing.

5.4. Navigating Biological Complexity and the “Black Box” Problem

The intrinsic complexity of genotype-phenotype relationships poses significant challenges for crop genetics. Traditional linear models underlying GWAS fail to capture epistasis, polygenic interactions, and G × E interactions (Wang et al., 2025c). While nonlinear deep learning architectures (CNNs, LSTMs) outperform linear models in predicting complex traits, their "black box" nature limits biological interpretability and practical breeding utility (Zhou et al., 2024). Integrating explainable AI (XAI) approaches like SHAP with frameworks such as G2PDiffusion offers a promising path to balance predictive accuracy with mechanistic insight (Figure 3). Complementary strategies, including dimensionality reduction (PCA, autoencoders), sparse modeling (LASSO, elastic net), multi-task learning, meta-learning, and causal inference, further enhance model interpretability and robustness across environments (Montesinos-López et al., 2018; Montesinos-López et al., 2021; Ogutu et al., 2011; Vettoruzzo et al., 2024).

Optimized data utilization through augmentation, transfer learning from model plants, active learning, and privacy-preserving federated learning helps mitigate data scarcity (Zhu et al., 2023). Nevertheless, current AI technologies only partially address high dimensionality and sparse data challenges, with robust cross-environment prediction for polygenic traits remaining a critical obstacle. Future advances will require synergistic progress in algorithmic innovation and high-quality data resource development.

5.5. Bridging the Lab-to-Field Divide

Despite computational advances, the translation from *in silico* prediction to field-stable varieties faces a "last mile" bottleneck defined by data scarcity, biological complexity, and genetic background effects.

A prominent disparity exists in crop improvement between the accelerated advancement of GE-AI and the protracted timeline of conventional field validation. To expedite this final validation phase, a suite of technologies is being synergistically integrated, containing speed breeding (enables 4-6 generations annually under controlled conditions), compressing generational turnover; high-throughput phenomics encompassing drone-based remote sensing, fixed sensor arrays, and automated imaging platforms, to facilitate large-scale, automated monitoring. The optimized experimental designs coupled with AI-driven models enhance the prediction of field performance from preliminary data, thereby refining resource allocation. Furthermore, cellular-level screening techniques, such as single-cell sequencing and high-content imaging, allow for the prioritization of promising gene-editing candidates prior to plant regeneration or field trials. This is achieved by developing AI models that correlate cellular phenotypes with complex agronomic traits, substantially alleviating the burden of subsequent validation. Collectively, these innovations establish a closed-loop pipeline from "design-edit" to "validate-optimize," propelling the comprehensive acceleration of crop improvement.

5.5.1. The Data-Complexity Conundrum

For many orphan crops and complex agronomic traits, data scarcity remains a significant barrier. Breeding programs generate small, high-dimensional datasets that fail to capture the full spectrum of G × E × M (Genotype × Environment × Management) interactions, greatly increasing susceptibility to model overfitting (Han et al., 2025). While strategies such as biologically informed data augmentation, transfer learning from model plants, and federated learning (see Table 1) across institutions (Zhu *et al.*, 2023) offer partial remedies, they may also introduce new biases due to distributional shifts between species and artifacts from synthetic data. Current AI architectures remain fundamentally data-scarce, and for polygenic traits governed by epistatic networks that defy linear assumptions (Wang *et al.*, 2025c), robust cross-environment predictions remain elusive.

Explainable AI (XAI) (see Table 1) approaches, such as SHAP analyses, may improve interpretability (Zhou *et al.*, 2024), but they do not resolve the underlying paucity of ground-truth biological data.

5.5.2. From Genotype to Phenotype: The Background Effect

A more formidable challenge lies in variety-specific genetic backgrounds. The same edit can yield opposite phenotypes in different varieties due to epistatic interactions and chromatin context, rendering current "one-size-fits-all" prediction models brittle (Borevitz et al., 2014; Canver et al., 2018). Addressing this requires a shift toward context-aware precision editing through constructing integrated pan-genome, epigenome, and transcriptome atlases covering core germplasms (Hufford et al., 2021) to decipher variety-specific regulatory networks. Building upon this, developing "context-aware" predictive models is paramount. This encompasses employing transfer learning to integrate data from disparate varieties to enhance model generalizability, and constructing deep learning models that amalgamate gene networks with whole-genome variation profiles of varieties, enabling *in silico* simulation of potential outcomes from editing perturbations across different backgrounds. Concurrently, experimental strategies require synergistic optimization, such as conducting early-stage parallel validation in multiple representative varieties, and designing gRNAs based on variety-specific sequences or employing broad-spectrum Cas variants. The envisioned "digital twin", where a variety of multi-omics profile predicts optimal editor-target combinations before experimental deployment (Varshney et al., 2020), remains aspirational, contingent on global data-sharing frameworks.

5.5.3. Regulatory Fragmentation Hurdle

Regulatory fragmentation further complicates this situation, bridging the lab-to-field divide on this planet (Figure 3). Policies governing New Genomic Techniques (NGTs) vary widely across jurisdictions: the European Union generally classifies gene-edited crops as GMOs, the U.S. USDA exempts edits lacking foreign DNA, and countries such as Argentina and Japan follow product-based, case-by-case frameworks (Entine et al., 2021; Jones et al., 2022). This patchwork of regulatory positions creates significant market uncertainty, discouraging the investment required for extensive field trials and reinforcing a negative feedback loop: the less regulatory clarity, the fewer field validations occur, and the sparser the data available to inform policy.

Collectively, these hurdles underscore that GE-AI synergy is not merely a computational problem but a systems integration challenge. Overcoming the lab-to-field gap will require not just algorithmic innovation, but coordinated advances in robotic automation, multi-omics data infrastructure, and harmonized governance frameworks.

6. Future Priorities: Toward Equitable and Scalable Solutions

Realizing the full potential of GE-AI integration will require advances that simultaneously expand computational capability and enable biological scalability (Figure 3). Key priorities are listed below.

6.1. Next-Generation AI Architectures and Design Paradigms

First, federated learning and explainable AI (XAI) must become a priority to enable collaborative modeling across proprietary germplasm collections while rendering AI decisions interpretable for regulatory trust (Table 1). Second, the shift from single-gene editing to systems-level design requires pan-genomic atlases that map epistatic interactions across diverse genetic backgrounds (Benoit et al., 2025), addressing the "variety-specific variability" challenge (Section 5.5.2). Large language models and generative AI (e.g., CRISPR-GPT) will automate experimental design, lowering technical barriers for non-specialists (Qu *et al.*, 2025). Third, chromosome-scale engineering leveraging emerging Cas-derived "molecular clamps" for megabase-level rearrangements represents the next frontier, demanding AI architectures capable of modeling three-dimensional genome architecture and predicting the phenotypic outcomes of large-scale structural variations.

6.2. Regulatory Harmonization and Global Convergence

Current regulatory fragmentation creates market uncertainty that stifles field validation. While the United States and Canada employ product-based frameworks (focusing on novel characteristics and foreign DNA absence rather than the editing process), the European Union maintains process-based GMO classification, though recent reforms signal movement toward proportionate oversight. Bridging this divide requires international alignment through bodies such as FAO and OECD on risk classification standards and mutual recognition of safety data. Crucially, regulatory frameworks must evolve to evaluate AI-designed organisms, where the "black box" nature of generative models challenges traditional risk assessment protocols.

6.3. Equitable Access and Public Trust

The GE-AI pipeline's data intensity risks exacerbating the "digital divide" between well-resourced institutions and smallholder farmers, particularly in the Global South. Governance must safeguard farmer data sovereignty while ensuring open-source toolkits remain accessible. Public acceptance hinges on transparent communication and rigorous impact assessment—not merely technical safety, but socio-economic and environmental equity (Ali et al., 2024). Only through interdisciplinary collaboration spanning plant biologists, computational scientists, ethicists, and policymakers can these technologies deliver resilient, sustainable, and just agricultural innovation.

7. Concluding Remarks

The convergence of GE and AI is beginning to shift crop improvement from a largely experience-driven craft into a predictive, programmable discipline. Rather than representing a fully mature paradigm, this convergence is most convincingly emerging in predictive design, model-guided optimization, and iterative experimental validation. Together, these technologies function as a reciprocal innovation cycle: AI identifies optimal genomic targets and editing strategies at scale, while GE executes these designs with molecular precision. In favorable cases, this partnership may accelerate design–build–test cycles and facilitate the more efficient combination of complex traits such as yield, disease resistance, and nutritional quality.

Throughout this perspective, we have highlighted how AI can serve as a computational engine that accelerates every stage of genome editing, from *de novo* protein and editor design to guide RNA optimization, editing-outcome prediction, and off-target risk reduction. Conversely, GE provides the biological validation framework essential for grounding AI models in empirical reality. The high-resolution, multi-omics datasets generated through GE feed directly back into AI pipelines, refining predictive accuracy and enabling continuous model improvement (Figure 1). This iterative synergy is becoming increasingly visible in applications related to climate and biotic stress resilience, nutritional enhancement, and *de novo* domestication, even though its broader implementation in routine crop breeding remains at an early stage (Figure 2).

A major next step for the field is to move from promising but fragmented tools toward benchmarked and context-aware predictive frameworks. Without harmonized reference datasets, comparable evaluation metrics, and systematic cross-species validation, it will remain difficult to distinguish robust and transferable AI-enabled editing models from study-specific performance claims. Community-endorsed standards for benchmarking and validation will therefore be essential for translating methodological advances into reliable crop genome-engineering platforms.

Looking ahead, the incorporation of real-time environmental sensing, autonomous editing systems, and generative biological design may further expand the development of climate-adaptive, resource-efficient crops. Yet the success of these scientific innovations will depend equally on their societal integration. Transparent and harmonized regulatory frameworks, equitable access to enabling technologies, and proactive public engagement are not optional add-ons but foundational components of responsible deployment.

Fully realizing the promise of GE-AI convergence will require sustained collaboration across disciplines, bringing together plant biologists, computational scientists, ethicists, social scientists, and policymakers. By aligning technological innovation with governance structures that prioritize safety, equity, and sustainability, we can harness this powerful synergy to address urgent global agricultural challenges and help secure a more resilient and food-secure future (Figure 3).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2023YFD1202900), the National Science Foundation of China (32372154), and the earmarked fund for CARS-02-06 to C. X.

Author contributions

C. X. conceived the overall structure, narrative of the perspective, and revised the full text and figures; Z. G. and J. Z. drafted major portions of the manuscript, designed and modified the figures, and participated in the revision of the manuscript. All authors approved the final version.

Declaration of Interests

The authors declare no conflicts of interest.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used AI-assisted technologies for language polishing and/or schematic figure refinement. The authors carefully reviewed and edited all outputs and take full responsibility for the content of the publication.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature **630**:493-500. 10.1038/s41586-024-07487-w.

Ali, G., Mijwil, M.M., Bosco Apparatus, B., Abotaleb, M., and Adamopoulos, I. (2024). A survey on artificial intelligence in cybersecurity for smart agriculture: State-of-the-art, cyber threats, artificial intelligence applications, and ethical concerns. Mesopotamian Journal of Computer Science **2024**:53-103. 10.58496/mjcsc/2024/007.

Alipanahi, B., Delong, A., Weirauch, M.T., and Frey, B.J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology **33**:831-838. 10.1038/nbt.3300.

Altae-Tran, H., Kannan, S., Suberski, A.J., Mears, K.S., Demircioglu, F.E., Moeller, L., Kocalar, S., Oshiro, R., Makarova, K.S., Macrae, R.K., et al. (2023). Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. Science **382**:1-14. 10.1126/science.adi1910.

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P., and Kelley, D.R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods **18**:1196-1203. 10.1038/s41592-021-01252-x.

Azoitei, M.L., Correia, B.E., Ban, Y.-E.A., Carrico, C., Kalyuzhniy, O., Chen, L., Schroeter, A., Huang, P.-S., McLellan, J.S., Kwong, P.D., et al. (2011). Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. Science **334**:373-376.

Benoit, M., Jenike, K.M., Satterlee, J.W., Ramakrishnan, S., Gentile, I., Hendelman, A., Passalacqua, M.J., Suresh, H., Shohat, H., Robitaille, G.M., et al. (2025). Solanum pan-genetics reveals paralogues as contingencies in crop engineering. Nature **640**:135-145. 10.1038/s41586-025-

08619-6.

Berman, A., Su, N., Li, Z., Landau, U., Chakraborty, J., Gerbi, N., Liu, J., Qin, Y., Yuan, B., Wei, W., et al. (2025). Construction of multi-targeted CRISPR libraries in tomato to overcome functional redundancy at genome-scale level. Nature Communications **16**:1-14. 10.1038/s41467-025-59280-6.

Borevitz, J.O., Wallace, J.G., Bradbury, P.J., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E.S. (2014). Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize. PLoS Genetics **10**:1-10. 10.1371/journal.pgen.1004845.

Budisa, N., Pak, M.A., Markhieva, K.A., Novikova, M.S., Petrov, D.S., Vorobyev, I.S., Maksimova, E.S., Kondrashov, F.A., and Ivankov, D.N. (2023). Using AlphaFold to predict the impact of single mutations on protein stability and function. Plos One **18**:1-9. 10.1371/journal.pone.0282689.

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M.V., and Komar, A.A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Molecular Cell **61**:341-351. 10.1016/j.molcel.2016.01.008.

Canver, M.C., Joung, J.K., and Pinello, L. (2018). Impact of Genetic Variation on CRISPR-Cas Targeting. The CRISPR Journal **1**:159-170. 10.1089/crispr.2017.0016.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. Annual Review of Plant Biology **70**:667-697. 10.1146/annurev-arplant-050718-100049.

Chen, Q., Chuai, G., Zhang, H., Tang, J., Duan, L., Guan, H., Li, W., Li, W., Wen, J., Zuo, E., et al. (2023). Genome-wide CRISPR off-target prediction and optimization using RNA-DNA interaction fingerprints. Nature Communications **14:7521**:1-17. 10.1038/s41467-023-42695-4.

Chen, X., Cui, X., and Ma, H. (2025). The application of artificial intelligence and gene editing technologies in high-yield crop breeding: Innovations and prospects for future agriculture. Advances in Resources Research **pp**:146-166. 10.50908/arr.5.1_146.

Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B., et al. (2018). DeepCRISPR: optimized CRISPR guide RNA design by deep learning. Genome Biology **19:80**:1-18. 10.1186/s13059-018-1459-4.

Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Research **46**:W242-W245. 10.1093/nar/gky354.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., Haas, R.J.d., Bethel, N., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. Science **378**:49-56.

Dixit, S., Kumar, A., Srinivasan, K., Vincent, P.M.D.R., and Ramu Krishnan, N. (2024). Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions. Frontiers in Bioengineering and Biotechnology **11**:1-16. 10.3389/fbioe.2023.1335901.

Entine, J., Felipe, M.S.S., Groenewald, J.-H., Kershen, D.L., Lema, M., McHughen, A., Nepomuceno, A.L., Ohsawa, R., Ordonio, R.L., Parrott, W.A., et al. (2021). Regulatory approaches for genome edited agricultural plants in select countries and jurisdictions around the world. Transgenic Research **30**:551-584. 10.1007/s11248-021-00257-8.

Farooq, M.A., Gao, S., Hassan, M.A., Huang, Z., Rasheed, A., Hearne, S., Prasanna, B., Li, X., and Li, H. (2024). Artificial intelligence in plant breeding. Trends in Genetics **40**:891-908. 10.1016/j.tig.2024.07.001.

Fei, H., Li, Y., Liu, Y., Wei, J., Chen, A., and Gao, C. (2025). Advancing protein evolution with inverse folding models integrating structural and evolutionary constraints. Cell **188**:1–19.

10.1016/j.cell.2025.06.014.

Fernie, A.R., and Yan, J. (2019). De novo domestication: An alternative route toward new crops for the future. Molecular Plant **12**:615-631. 10.1016/j.molp.2019.03.016.

Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology **31**:397-405. 10.1016/j.tibtech.2013.04.004.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. Cell **184**:1621-1635. 10.1016/j.cell.2021.01.005.

Gao, P., Lian, L., Feng, W., Ma, Y., Lin, J., Qin, L., Hao, S., Zhao, H., Liu, X., Yuan, J., et al. (2025). Deciphering the sequence basis and application of transcriptional initiation regulation in plant genomes through deep learning. Genome Biology **26**:1-26. 10.1186/s13059-025-03782-5.

Han, J., Shi, L., Yang, Q., Yu, J., and Athanasiadis, I.N. (2025). Knowledge-guided machine learning with multivariate sparse data for crop growth modelling. Field Crops Research **328**:1-25. 10.1016/j.fcr.2025.109912.

Huang, J., Lin, Q., Fei, H., He, Z., Xu, H., Li, Y., Qu, K., Han, P., Gao, Q., Li, B., et al. (2023). Discovery of deaminase functions by structure-based protein clustering. Cell **186**:3182-3195.e3114. 10.1016/j.cell.2023.05.041.

Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Chougule, K.M., Ou, S., Liu, J., Ricci, W.A., Guo, T., Olson, A., Qiu, Y., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science **373**:655-662.

Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., Zhang, J., Ness, T.J., Banjade, S., Bagde, S.R., et al. (2021). Computed structures of core eukaryotic protein complexes. Science **374**:1-12. 10.1126/science.abm4805.

Ichikawa, D.M., Abdin, O., Alerasool, N., Kogenaru, M., Mueller, A.L., Wen, H., Giganti, D.O., Goldberg, G.W., Adams, S., Spencer, J.M., et al. (2023). A universal deep-learning model for zinc finger design enables transcription factor reprogramming. Nature Biotechnology **41**:1117-1129. 10.1038/s41587-022-01624-4.

Jeon, Y.-J., Hong, S., Lee, T.S., Park, S.H., Song, G., Seo, M.-G., Lee, J., Lim, Y., An, J.-T., Lee, S., et al. (2025). Volumetric Deep Learning-Based Precision Phenotyping of Gene-Edited Tomato for Vertical Farming. Plant Phenomics **7**:1-15. 10.1016/j.plaphe.2025.100095.

Jiang, K., Yan, Z., Bernardo, M.D., Sgrizzi, S.R., Villiger, L., Kayabolen, A., Kim, B.J., Carscadden, J.K., Hiraizumi, M., Nishimasu, H., et al. (2024). Rapid in silico directed evolution by a protein language model with EVOLVEpro. Science **387(6732):eadr6006**:1-21. 10.1126/science.adr6006.

Jin, S., Zhu, Z., Li, Y., Zhang, S., Liu, Y., Li, D., Li, Y., Luo, Y., Cheng, Z., Zhao, K.T., et al. (2025). Functional RNA splitting drove the evolutionary emergence of type V CRISPR-Cas systems from transposons. Cell **188**:6283-6300.e6222. 10.1016/j.cell.2025.09.004.

Jones, M.G.K., Fosu-Nyarko, J., Iqbal, S., Adeel, M., Romero-Aldemita, R., Arujanan, M., Kasai, M., Wei, X., Prasetya, B., Nugroho, S., et al. (2022). Enabling trade in gene-edited produce in Asia and Australasia: The developing regulatory landscape and future perspectives. Plants **11**:1-36. 10.3390/plants11192538.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature **596**:583-589. 10.1038/s41586-021-03819-2.

Kim, N., Choi, S., Kim, S., Song, M., Seo, J.H., Min, S., Park, J., Cho, S.-R., and Kim, H.H. (2023). Deep learning models to predict the editing efficiencies and outcomes of diverse base editors. Nature

Biotechnology **42**:484-497. 10.1038/s41587-023-01792-x.

Kompas, T., Che, T.N., and Grafton, R.Q. (2024). Global impacts of heat and water stress on food production and severe food insecurity. Scientific Reports **14**:1-9. 10.1038/s41598-024-65274-z.

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Research **47**:W171-W174. 10.1093/nar/gkz365.

Lei, Y., Lu, L., Liu, H.-Y., Li, S., Xing, F., and Chen, L.-L. (2014). CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants. Molecular Plant **7**:1494-1496. 10.1093/mp/ssu044.

Lenaerts, B., Collard, B.C.Y., and Demont, M. (2019). Review: Improving global food security through accelerated plant breeding. Plant Science **287**:1-8. 10.1016/j.plantsci.2019.110207.

Li, F., Tan, X., Li, S., Chen, S., Liu, L., Huang, J., Li, G., Lu, Z., Wu, J., Zeng, D., et al. (2025a). SuperDecode: An integrated toolkit for analyzing mutations induced by genome editing. Molecular Plant **18**:690-702. 10.1016/j.molp.2025.03.002.

Li, G., An, L., Yang, W., Yang, L., Wei, T., Shi, J., Wang, J., Doonan, J.H., Xie, K., Fernie, A.R., et al. (2025b). Integrated biotechnological and AI innovations for crop improvement. Nature **643**:925-937. 10.1038/s41586-025-09122-8.

Li, S., An, J., Li, Y., Zhu, X., Zhao, D., Wang, L., Sun, Y., Yang, Y., Bi, C., Zhang, X., et al. (2022a). Automated high-throughput genome editing platform with an AI learning in situ prediction model. Nature Communications **13**:1-11. 10.1038/s41467-022-35056-0.

Li, S., Lin, D., Zhang, Y., Deng, M., Chen, Y., Lv, B., Li, B., Lei, Y., Wang, Y., Zhao, L., et al. (2022b). Genome-edited powdery mildew resistance in wheat without growth penalties. Nature **602**:455-460. 10.1038/s41586-022-04395-9.

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., Dong, W., Gao, C., and Xu, C. (2018). Domestication of wild tomato is accelerated by genome editing. Nature Biotechnology **36**:1160-1163. 10.1038/nbt.4273.

Li, W., Jiang, X., Wang, W., Hou, L., Cai, R., Li, Y., Gu, Q., Chen, Q., Ma, P., Tang, J., et al. (2024). Discovering CRISPR-Cas system with self-processing pre-crRNA capability by foundation models. Nature Communications **15**:1-14. 10.1038/s41467-024-54365-0.

Lin, Q., Jin, S., Zong, Y., Yu, H., Zhu, Z., Liu, G., Kou, L., Wang, Y., Qiu, J.-L., Li, J., et al. (2021). High-efficiency prime editing with optimized, paired pegRNAs in plants. Nature Biotechnology **39**:923-927. 10.1038/s41587-021-00868-w.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, O., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**:1123–1130. 10.1126/science.ade2574.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., and Chen, L.-L. (2017). CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. Molecular Plant **10**:530-532. 10.1016/j.molp.2017.01.003.

Lorenzo, C.D., Debray, K., Herwegh, D., Develtere, W., Impens, L., Schaumont, D., Vandeputte, W., Aesaert, S., Coussens, G., De Boe, Y., et al. (2023). BREEDIT: a multiplex genome editing strategy to improve complex quantitative traits in maize. The Plant Cell **35**:218-238. 10.1093/plcell/koac243.

Lou, H., Li, S., Shi, Z., Zou, Y., Zhang, Y., Huang, X., Yang, D., Yang, Y., Li, Z., and Xu, C. (2025). Engineering source-sink relations by prime editing confers heat-stress resilience in tomato and rice. Cell **188**:530-549.e520. 10.1016/j.cell.2024.11.005.

Maas, S., Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. Plos One **10**:1-11. 10.1371/journal.pone.0124633.

Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos, J.L., Xiong, C., Sun, Z.Z., Socher, R., et al. (2023). Large language models generate functional protein sequences across diverse families. Nature Biotechnology **41**:1099-1106. 10.1038/s41587-022-01618-2.

Mathis, N., Allam, A., Tálas, A., Kissling, L., Benvenuto, E., Schmidheini, L., Schep, R., Damodharan, T., Balázs, Z., Janjuha, S., et al. (2024). Machine learning prediction of prime editing efficiency across diverse chromatin contexts. Nature Biotechnology **43**:712-719. 10.1038/s41587-024-02268-2.

Montesinos-López, O.A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C.M., and Martín-Vallejo, J. (2018). Multi-trait, Multi-environment Deep Learning Modeling for Genomic-Enabled Prediction of Plant Traits. G3 Genes|Genomes|Genetics **8**:3829-3840. 10.1534/g3.118.200728.

Montesinos-López, O.A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J.A., Martini, J.W.R., Fajardo-Flores, S.B., Gaytan-Lugo, L.S., Santana-Mancilla, P.C., and Crossa, J. (2021). A review of deep learning applications for genomic selection. BMC Genomics **22**:1-23. 10.1186/s12864-020-07319-x.

Ngou, B.P.M., Wyler, M., Schmid, M.W., Suzuki, T., Albert, M., Dohmae, N., Kadota, Y., and Shirasu, K. (2025). Systematic discovery and engineering of synthetic immune receptors in plants. Science **389**:1-14. 10.1126/science.adx2508.

Nguyen, E., Poli, M., Durrant, M.G., Kang, B., Katrekar, D., Li, D.B., Bartie, L.J., Thomas, A.W., King, S.H., Brixi, G., et al. (2024). Sequence modeling and design from molecular to genome scale with Evo. Science **386**:1-21. 10.1126/science.ado9336.

Notin, P., Rollins, N., Gal, Y., Sander, C., and Marks, D. (2024). Machine learning for functional protein design. Nature Biotechnology **42**:216-228. 10.1038/s41587-024-02127-0.

Ogutu, J.O., Piepho, H.-P., and Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. BMC Proceedings **5**:1-5.

Park, J., Bae, S., and Kim, J.-S. (2015). Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics **31**:4014-4016. 10.1093/bioinformatics/btv537.

Pei, Y., Zhao, Y., Wang, H., Guo, Y., Gu, X., Lv, J., Guo, Z., Chen, Y., Ren, Y., Ren, Y., et al. (2025). Uncovering Convergent Pattern Recognition Receptors Recognising Phytophthora Across Plant Lineages. Plant Biotechnology Journal:1-12. 10.1111/pbi.70409.

Qiu, Y., Liu, L., Yan, J., Xiang, X., Wang, S., Luo, Y., Deng, K., Xu, J., Jin, M., Wu, X., et al. (2025). Precise engineering of gene expression by editing plasticity. Genome Biology **26**:1-28. 10.1186/s13059-025-03516-7.

Qu, Y., Huang, K., Yin, M., Zhan, K., Liu, D., Yin, D., Cousins, H.C., Johnson, W.A., Wang, X., Shah, M., et al. (2025). CRISPR-GPT for agentic automation of gene-editing experiments. Nature Biomedical Engineering:1-19. 10.1038/s41551-025-01463-z.

Richards, B., Tsao, D., and Zador, A. (2022). The application of artificial intelligence to biology and neuroscience. Cell **185**:2640-2643. 10.1016/j.cell.2022.06.047.

Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S., et al. (2019). A deep learning framework for neuroscience. Nature Neuroscience **22**:1761-1770. 10.1038/s41593-019-0520-2.

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C.L., Ma, J., et

al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences **118**:1-12. 10.1073/pnas.2016239118.

Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E., and Lippman, Z.B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. Cell **171**:470-480.e478. 10.1016/j.cell.2017.08.030.

Ruffolo, J.A., Nayfach, S., Gallagher, J., Bhatnagar, A., Beazer, J., Hussain, R., Russ, J., Yip, J., Hill, E., Pacesa, M., et al. (2025). Design of highly functional genome editors by modelling CRISPR–Cas sequences. Nature **645**:518-525. 10.1038/s41586-025-09298-z.

Sha, G., Sun, P., Kong, X., Han, X., Sun, Q., Fouillen, L., Zhao, J., Li, Y., Yang, L., Wang, Y., et al. (2023). Genome editing of a rice CDP-DAG synthase confers multipathogen resistance. Nature **618**:1017-1023. 10.1038/s41586-023-06205-2.

Silverstein, R.A., Kim, N., Kroell, A.-S., Walton, R.T., Delano, J., Butcher, R.M., Pacesa, M., Smith, B.K., Christie, K.A., Ha, L.L., et al. (2025). Custom CRISPR–Cas9 PAM variants via scalable engineering and machine learning. Nature **643**:539-550. 10.1038/s41586-025-09021-y.

Song, X., Meng, X., Guo, H., Cheng, Q., Jing, Y., Chen, M., Liu, G., Wang, B., Wang, Y., Li, J., et al. (2022). Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. Nature Biotechnology **40**:1403-1411. 10.1038/s41587-022-01281-7.

Starr, T.N., and Thornton, J.W. (2016). Epistasis in protein evolution. Protein Science **25**:1204-1218. 10.1002/pro.2897.

Sulis, D.B., Jiang, X., Yang, C., Marques, B.M., Matthews, M.L., Miller, Z., Lan, K., Cofre-Vega, C., Liu, B., Sun, R., et al. (2023). Multiplex CRISPR editing of wood for sustainable fiber production. Science **381**:216–221. 10.1126/science.add4514.

Sun, C., Li, H., Liu, Y., Li, Y., Gao, R., Shi, X., Fei, H., Liu, J., Liang, R., and Gao, C. (2025). Iterative recombinase technologies for efficient and precise genome engineering across kilobase to megabase scales. Cell **188**:1–18. 10.1016/j.cell.2025.07.011.

Tianyi Li , Hui Xu, Shouzhen Teng, Mingrui Suo, Revocatus Bahitwa, Mingchi Xu, Yiheng Qian, Guillaume P. Ramstein, Baoxing Song, Edward S. Buckler, et al. (2024). Modeling 0.6 million genes for the rational design of functional cis-regulatory variants and de novo design of cis- regulatory sequences. Proceedings of the National Academy of Sciences **121**:1-9. 10.1073/pnas.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature **596**:590-596. 10.1038/s41586-021-03828-1.

van Dijk, A.D.J., Kootstra, G., Kruijer, W., and de Ridder, D. (2021). Machine learning in plant science and plant breeding. iScience **24**:1-12. 10.1016/j.isci.2020.101890.

Varshney, R.K., Sinha, P., Singh, V.K., Kumar, A., Zhang, Q., and Bennetzen, J.L. (2020). 5Gs for crop genetic improvement. Current Opinion in Plant Biology **56**:190-196. 10.1016/j.pbi.2019.12.004.

Vettoruzzo, A., Bouguelia, M.-R., Vanschoren, J., Rögnvaldsson, T., and Santosh, K.C. (2024). Advances and Challenges in Meta-Learning: A Technical Review. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**:4763-4779. 10.1109/tpami.2024.3357847.

Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., et al. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. Nature Communications **10**:1-14. 10.1038/s41467-019-12281-8.

Wang, H., Yan, S., Wang, W., Chen, Y., Hong, J., He, Q., Diao, X., Lin, Y., Chen, Y., Cao, Y., et al.

(2025a). Cropformer: An interpretable deep learning framework for crop genomic prediction. Plant Communications **6**:1-14. 10.1016/j.xplc.2024.101223.

Wang, J., Zhang, L., Wang, S., Wang, X., Li, S., Gong, P., Bai, M., Paul, A., Tvedt, N., Ren, H., et al. (2025b). AlphaFold‐Guided Bespoke Gene Editing Enhances Field‐Grown Soybean Oil Contents. Advanced Science **12**:1-11. 10.1002/advs.202500290.

Wang, J., Lisanza, S., Juergens, D., Tischer, D., Watson, J.L., Castro, K.M., Ragotte, R., Saragovi, A., Milles, L.F., Baek, M., et al. (2022). Scaffolding protein functional sites using deep learning. Science **377**:387–394. 10.1126/science.abn2100.

Wang, P.-Y., Chen, Z.-S., Jiao, X., and Bao, Y.-J. (2025c). Integrating deep learning models with genome-wide association study-based identification enhanced phenotype predictions in Group A Streptococcus. Journal of Microbiology and Biotechnology **35**:1-10. 10.4014/jmb.2411.11010.

Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature **620**:1089-1100. 10.1038/s41586-023-06415-8.

Wicky, B.I.M., Milles, L.F., Courbet, A., Ragotte, A., Dauparas, J., Kinfu, E., Tipps, S., Kibler, R.D., Baek, M., DiMaio, F., et al. (2022). Hallucinating symmetric protein assemblies. Science **378**:56-61.

Winnifrith, A., Outeiral, C., and Hie, B.L. (2024). Generative artificial intelligence for de novo protein design. Current Opinion in Structural Biology **86**:1-16. 10.1016/j.sbi.2024.102794.

Wu, K., Jiang, H., Hicks, D.R., Liu, C., Muratspahić, E., Ramelot, T.A., Liu, Y., McNally, K., Kenny, S., Mihut, A., et al. (2025). Design of intrinsically disordered region binding proteins. Science **389**:1-9. 10.1126/science.adr8063.

Wu, Y., He, Y., Sretenovic, S., Liu, S., Cheng, Y., Han, Y., Liu, G., Bao, Y., Fang, Q., Zheng, X., et al. (2021). CRISPR‐BETS: a base‐editing design tool for generating stop codons. Plant Biotechnology Journal **20**:499-510. 10.1111/pbi.13732.

Xia, Y., Liang, Z., Du, X., Cao, D., Li, J., Sun, L., Huo, Y.-X., and Guo, S. (2025). Design of function-regulating RNA via deep learning and AlphaFold 3. Briefings in Bioinformatics **26**:1-12. 10.1093/bib/bbaf419.

Xiao, Z., Li, W., Moon, H., Roell, G.W., Chen, Y., and Tang, Y.J. (2023). Generative artificial intelligence GPT-4 accelerates knowledge mining and machine learning for synthetic biology. ACS Synthetic Biology **12**:2973-2982. 10.1021/acssynbio.3c00310.

Xie, T., and Zhang, Z. (2025). Construction of large-scale maize drought resistance gene networks based on multi-omics integration and identification of keyregulatory factors. Advances in Resources Research **pp**:203-234. 10.50908/arr.5.1_203.

Xie, X., Ma, X., Zhu, Q., Zeng, D., Li, G., and Liu, Y.-G. (2017). CRISPR-GE: A Convenient Software Toolkit for CRISPR-Based Genome Editing. Molecular Plant **10**:1246-1249. 10.1016/j.molp.2017.06.004.

Xu, J.-J., Lei, Y., Zhang, X.-F., Li, J.-X., Lin, Q., Wu, X.-D., Jiang, Y.-G., Zhang, W., Qian, R., Xiong, S.-Y., et al. (2025). Design of CoQ10 crops based on evolutionary history. Cell **188**:1941-1954.e1915. 10.1016/j.cell.2025.01.023.

Yao, Y., Chen, F., Wu, C., Chang, X., Cheng, W., Wang, Q., Deng, Z., Liu, T., and Lu, L. (2025). Structure‐based virtual screening aids the identification of glycosyltransferases in the biosynthesis of salidroside. Plant Biotechnology Journal **23**:1725-1735. 10.1111/pbi.70002.

Yoon, P.H., Zhang, Z., Loi, K.J., Adler, B.A., Lahiri, A., Vohra, K., Shi, H., Rabelo, D.B., Trinidad, M., Boger, R.S., et al. (2024). Structure-guided discovery of ancestral CRISPR-Cas13 ribonucleases.

Science **385**:538-543. 10.1126/science.adq0553.

Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., Chen, M., Jing, Y., Kou, L., Li, X., et al. (2021). A route to de novo domestication of wild allotetraploid rice. Cell **184**:1156-1170.e1114. 10.1016/j.cell.2021.01.013.

Yuan, X., Li, S., Chen, J., Yu, H., Yang, T., Wang, C., Huang, S., Chen, H., and Ao, X. (2024). Impacts of global climate change on agricultural production: A comprehensive review. Agronomy **14**:1-19. 10.3390/agronomy14071360.

Zambaldi, V., La, D., Chu, A.E., Patani, H., Danson, A.E., Kwan, T.O.C., Frerix, T., Schneider, R.G., Saxton, D., Thillaisundaram, A., et al. (2024). De novo design of high-affinity protein binders with AlphaProteo. arXiv:1-45.

Zhang, D., Xu, F., Wang, F., Le, L., and Pu, L. (2025a). Synthetic biology and artificial intelligence in crop improvement. Plant Communications **6**:1-18. 10.1016/j.xplc.2024.101220.

Zhang, J., Yu, H., and Li, J. (2023a). De novo domestication: retrace the history of agriculture to design future crops. Current Opinion in Biotechnology **81**10.1016/j.copbio.2023.102946.

Zhang, Y., Zhang, N., Chai, X., Sun, T., and Lawson, T. (2023b). Machine learning for image-based multi-omics analysis of leaf veins. Journal of Experimental Botany **74**:4928-4941. 10.1093/jxb/erad251.

Zhang, Y., Ikram, M., Khan, N., Zhao, N., Derakhshani, B., Usman, B., and Wang, H.-F. (2025b). Machine learning and functional validation identify OsRAV11/12 as negative regulators of drought tolerance and early flowering in rice. International Journal of Biological Macromolecules **329**:1-12. 10.1016/j.ijbiomac.2025.147709.

Zhao, P., Yang, H., Sun, Y., Zhang, J., Gao, K., Wu, J., Zhu, C., Yin, C., Chen, X., Liu, Q., et al. (2025). Targeted MYC2 stabilization confers citrus Huanglongbing resistance. Science **388**:191–198. 10.1126/science.adq7203.

Zhou, W., Yan, Z., and Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. Scientific Reports **14**:1-16. 10.1038/s41598-024-55243-x.

Zhu, Z., Lin, K., Jain, A.K., and Zhou, J. (2023). Transfer Learning in Deep Reinforcement Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**:13344-13362. 10.1109/tpami.2023.3292075.

Zsögön, A., Čermák, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L., Voytas, D.F., Kudla, J., and Peres, L.E.P. (2018). De novo domestication of wild tomato using genome editing. Nature Biotechnology **36**:1211-1216. 10.1038/nbt.4272.

FIGURE LEGENDS

Figure 1. The Closed-Loop Engine: Synergizing AI and GE for Precision Crop Breeding.

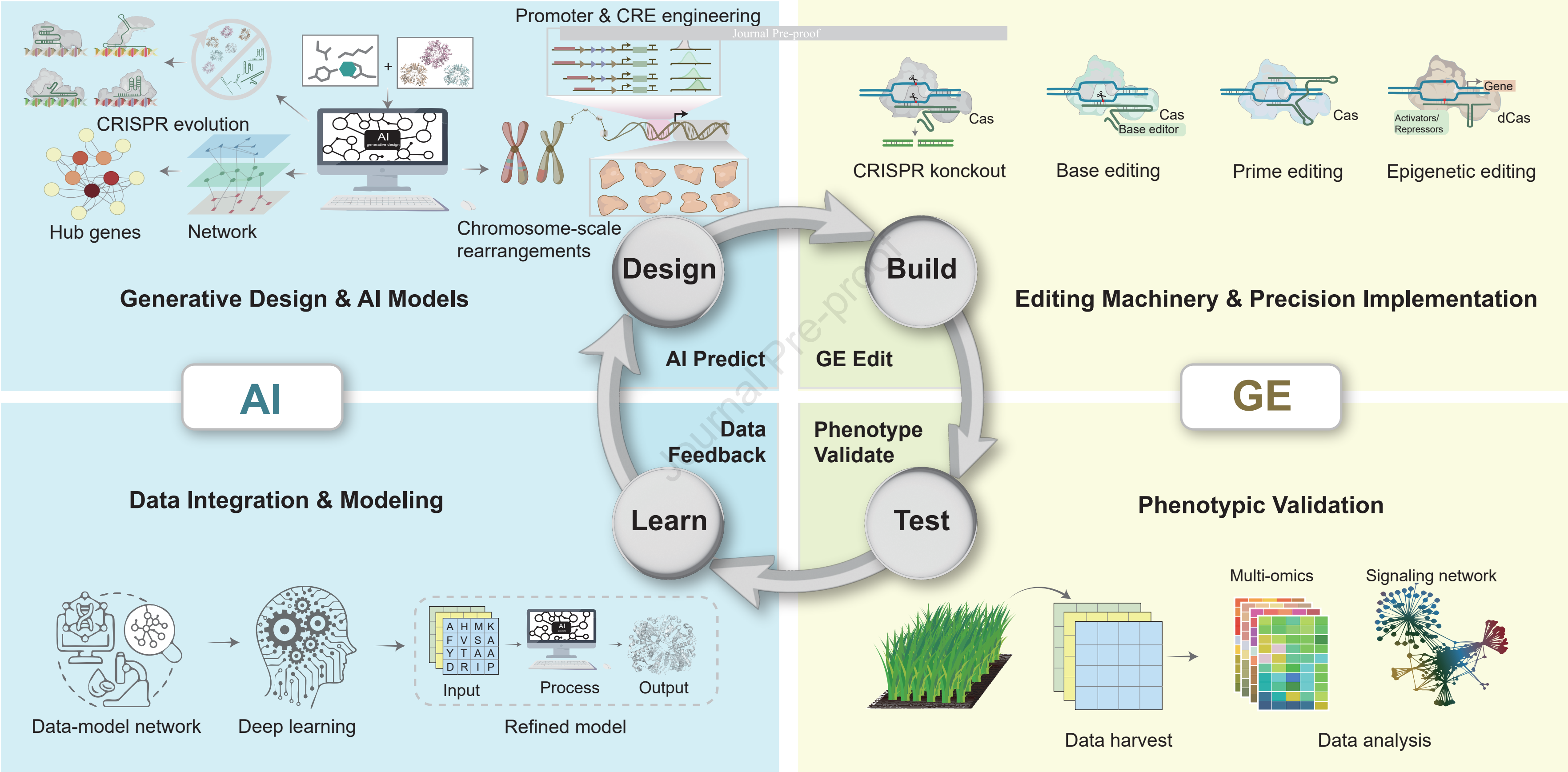
This schematic illustrates the bidirectional integration of Artificial Intelligence (AI) and Genome Editing (GE) through the Design-Build-Test-Learn (DBTL) framework. **The AI module** (left) comprises computational layers ***AI Predict*** (Generative Design and AI Models) and ***Data Feedback*** (Data integration & modeling). The **Generative Design** contains CRISPR evolution, sgRNA optimization for diverse editing modalities, *de novo* enzyme, multi-omics target prioritization (network & hub genes), regulatory element engineering (promoter & CRE engineering), and chromosome-scale rearrangements. The **AI Models** contains CNNs (applicable to sgRNA efficiency, motif detection), Transformers (applicable to promoter design, chromatin context), and Hybrid (applicable to enzyme design). The **Data integration & modeling** contain protein model refinement, multi-omics fusion, and G×E×M modeling. **The GE module *(right)*** constitutes the experimental execution layer: ***GE Edit*** (deployment of CRISPR/Cas systems and precision editing machinery) and ***Phenotype Validate*** (high-throughput phenotyping, field trials, and multi-omics data generation). The central DBTL engine depicts the iterative flow wherein AI-guided designs inform GE implementation, and empirical outcomes feedback to refine predictive models, creating a self-improving cycle for crop improvement.

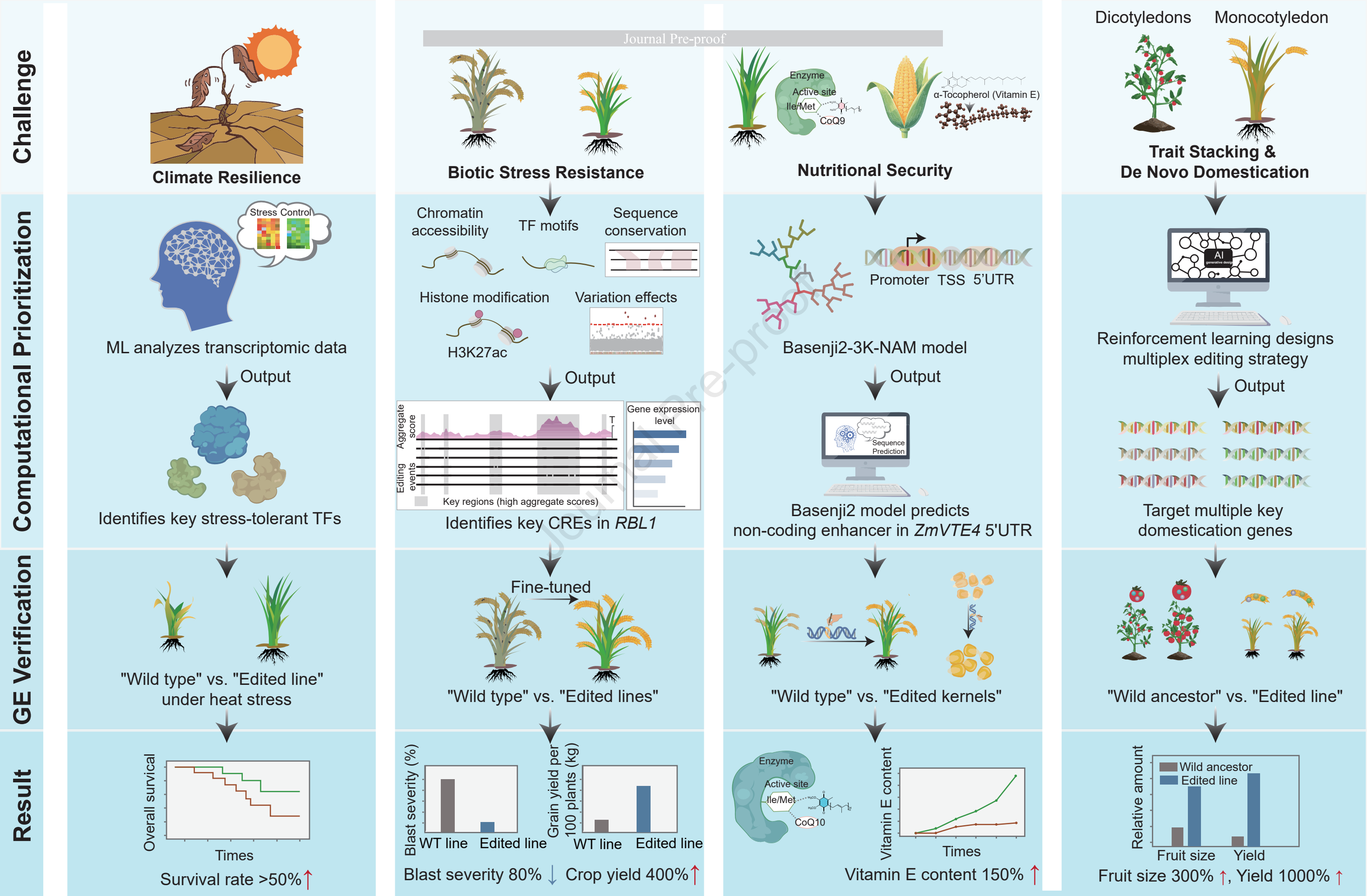
Figure 2. From *In Silico* Prediction to *In Planta* Validation: Case Studies of Emerging GE-AI Synergy.

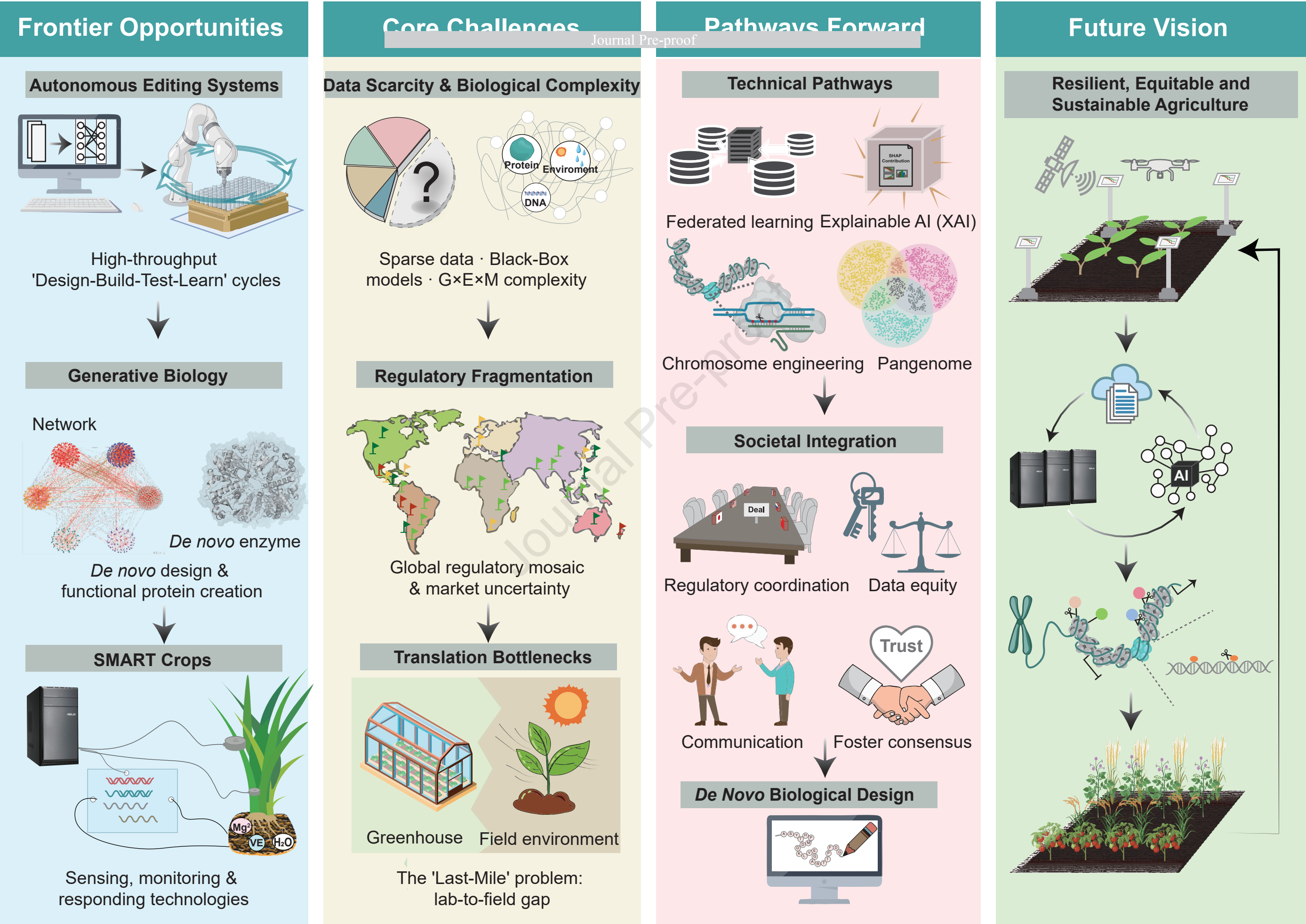
This figure illustrates the representative case studies depicting the pipeline from computational prediction to GE experimental validation. Each column traces a distinct challenge domain—**Climate Resilience**, **Biotic Stress Resistance**, **Nutritional Security**, and **Trait Stacking & *De Novo* Domestication**—through three sequential stages: ***Computational prioritization*** (top), *GE verification* (middle), and *phenotypic outcome* (bottom). ***Computational prioritization*** strategies include machine learning analysis of transcriptomic data to identify stress-tolerant transcription factors, chromatin accessibility profiling to discover cis-regulatory elements (e.g., in *RBL1*), the *Basenji2* model to predict non-coding enhancers (e.g., *ZmVTE4* 5'UTR), and reinforcement learning to design multiplex editing strategies for domestication genes. ***GE verification*** demonstrates precise implementation through prime editing, base editing, or multiplex CRISPR, validated under controlled conditions or field trials. ***Results (phenotypic outcome)*** quantify agronomic improvements, including enhanced survival rates under heat stress, 80% reduction in blast severity, 150% increase in vitamin E content, and a simultaneous improvement of fruit size and yield in domesticated lines. Vertical arrows indicate the directional flow from computational design to biological realization.

Figure 3. Navigating the Future: Opportunities, Hurdles, and Pathways for GE-AI Integration.

This framework depicts the forward-looking landscape for GE-AI technologies in agriculture, organized into four interconnected themes. ***Frontiers Opportunities*** (Column 1): Showcases transformative opportunities such as autonomous editing platforms, generative biological design, and SMART crops. ***Core Challenges*** (Column 2): Details major technical and translational obstacles, including data scarcity, biological complexity, fragmented regulatory environments, and the 'Last-Mile' problem. ***Pathways Forward*** (Column 3): Outlines key technological, societal integration, and methodological solutions, ranging from advanced AI architectures to *de novo* biological system design. ***Future Vision*** (Column 4): Converges on the overarching goal of achieving resilient, equitable, and sustainable agriculture, exemplified by the development of SMART crops: plants equipped with sensing, monitoring, and responding technologies. The left-to-right flow represents the progression from identifying opportunities and constraints to developing integrative solutions to realizing a future-ready agricultural system.


Promoter & CRE engineering
CRISPR evolution
AI
generative design
Hub genes
Network
Chromosome-scale rearrangements
Design
Generative Design & AI Models
AI Predict
AI
Data Feedback
Data Integration & Modeling
Learn
Data-model network
Deep learning
A H M K
F V S A
Y T A A
D R I P
Input
Process
Output
Refined model
Cas
CRISPR konckout
Cas
Base editor
Base editing
Cas
Prime editing
Gene
Activators/ Repressors
dCas
Epigenetic editing
Build
Editing Machinery & Precision Implementation
GE Edit
GE
Phenotype Validate
Test
Phenotypic Validation
Multi-omics
Signaling network
Data harvest
Data analysis


Challenge
Computational Prioritization
GE Verification
Result
Climate Resilience
Biotic Stress Resistance
Enzyme
Active site
Ile/Met
CoQ9
α-Tocopherol (Vitamin E)
Nutritional Security
Dicotyledons
Monocotyledon
Trait Stacking &
De Novo Domestication
Stress Control
ML analyzes transcriptomic data
Output
Identifies key stress-tolerant TFs
Chromatin accessibility
TF motifs
Sequence conservation
Histone modification
H3K27ac
Variation effects
Output
Aggregate score
Editing events
Key regions (high aggregate scores)
Gene expression level
Identifies key CREs in *RBL1*
Promoter
TSS
5'UTR
Basenji2-3K-NAM model
Output
Sequence Prediction
Basenji2 model predicts
non-coding enhancer in *ZmVTE4* 5'UTR
AI
Reinforcement learning designs
multiplex editing strategy
Output
Target multiple key
domestication genes
"Wild type" vs. "Edited line"
under heat stress
Fine-tuned
"Wild type" vs. "Edited lines"
"Wild type" vs. "Edited kernels"
"Wild ancestor" vs. "Edited line"
Overall survical
Times
Survival rate >50%↑
Blast severity (%)
WT line
Edited line
Grain yield per 100 plants (kg)
WT line
Edited line
Blast severity 80%↓
Crop yield 400%↑
Enzyme
Active site
Ile/Met
CoQ10
Vitamin E content
Times
Vitamin E content 150%↑
Relative amount
Wild ancestor
Edited line
Fruit size
Yield
Fruit size 300%↑, Yield 1000%↑


Frontier Opportunities
Autonomous Editing Systems
High-throughput 'Design-Build-Test-Learn' cycles
Generative Biology
Network
De novo enzyme
De novo design & functional protein creation
SMART Crops
Mg2+
VE
H2O
Sensing, monitoring & responding technologies
Core Challenges
Data Scarcity & Biological Complexity
?
Protein
Enviroment
DNA
Sparse data · Black-Box models · G×E×M complexity
Regulatory Fragmentation
Global regulatory mosaic & market uncertainty
Translation Bottlenecks
Greenhouse
Field environment
The 'Last-Mile' problem: lab-to-field gap
Pathways Forward
Technical Pathways
SHAP Contribution
Federated learning
Explainable AI (XAI)
Chromosome engineering
Pangenome
Societal Integration
Deal
Regulatory coordination
Data equity
Communication
Trust
Foster consensus
De Novo Biological Design
Future Vision
Resilient, Equitable and Sustainable Agriculture
AI