人工智能赋能现代农业生物育种

徐启玉[1]，郑 波[1,2*]，钟上威[1,3]

（1. 北京大学现代农业研究院，山东潍坊 261325，中国；2. 潍坊市科技创新促进中心，山东潍坊 261071，中国；3. 北京大学生命科学学院，北京 100871，中国）

摘 要：［**目的/意义**］随着基因型–表型–环境互作复杂性的增加及组学大数据的爆发式增长，传统育种已难以满足高效精准育种需求。驱动作物育种从经验选择向智能设计转型，对于保障全球粮食安全、实现农业可持续发展具有重要战略意义。本文旨在系统探究人工智能（Artificial Intelligence, AI）驱动植物育种范式变革的底层逻辑，梳理智慧育种的代际演变及其在表型解析、基因组预测和分子设计等关键环节的技术进展，并深入探讨育种大模型对育种全流程的重构作用，以期为中国种业科技实现自立自强提供理论参考。［**进展**］本文系统回顾了作物育种的代际演变历程，深入探讨了AI技术在现代育种中的应用现状和前景。重点阐述了AI在种质资源高效挖掘、高通量表型精准解析、全基因组选择模型预测及分子设计育种等关键环节的技术逻辑与实证案例。同时，剖析了农业大模型在重构农业知识体系、实现多模态数据融合建模以及辅助精准育种决策中的核心作用，并总结了当前AI驱动下育种技术框架的最新研究成果。［**结论/展望**］针对当前面临的跨学科数据壁垒、深度学习算法的可解释性不足，以及复杂变动环境下的模型泛化能力弱等核心挑战进行了深入分析。未来应致力于构建标准化育种大数据共享平台，开发具有生物学启发性的可解释算法，并加强AI与合成生物学的深度融合，通过多学科交叉协同，最终实现育种全流程的数字化、智能化与自动化演进。

关键词：智慧育种；人工智能；育种大模型；多组学融合；高通量表型；全基因组选择；智能设计育种

中图分类号：S330;S24;TP29 **文献标识码：**A **文章编号：**SA202602016

引用格式：徐启玉, 郑波, 钟上威 . 人工智能赋能现代农业生物育种[J]. 智慧农业(中英文), 2026, 8(3): 67-84. DOI：10.12133/j.smartag.SA202602016

XU Qiyu, ZHENG Bo, ZHONG Shangwei. Artificial Intelligence Empowering Modern Agricultural Biological Breeding [J]. Smart Agriculture, 2026, 8(3): 67-84. DOI：10.12133/j.smartag.SA202602016 (in Chinese with English abstract)

0 引 言

粮食安全是国家发展基石，种业创新则是保证粮食安全的战略核心[1]。作物育种是应对全球粮食安全与气候挑战的核心途径，通过培育高产、优质、抗逆新品种，可显著提升农业生产效率与韧性，为保障国家粮食安全提供战略支撑[2,3]。作物育种本质上是聚合优异等位基因，实现最佳亲本组配的过程。然而，作物产量的形成不仅受微效多基因网络的精密调控，更表现出显著的基因型（Genotype, G）×环境（Environment, E）交互效应[4]。传统育种长期依赖大规模田间表型筛选，育种周期长、效率低[5,6]，极难在海量候选组合中精准锁定具备高产潜力的“超级组合”，且由于长期追求高产单一性状的选择，导致作物遗传基础狭窄，在面对新兴病虫害及极端气候时愈发脆弱[7,8]。

随着高通量测序与精准表型平台的迅速发展，育种领域也进入到了多组学大数据时代[9]。然而，数据量的爆发并未直接转化为育种效率的飞跃，当前种业面临的核心科学问题在于：如何从复杂的“基因型-表型”非线性关联中，实现对复杂性状的跨时空预测？传统生物统计学模型在处理高维、非线性关联和复杂异构数据时往往力不从心，难以捕捉跨组学的深层关联。在此背景下，人工智能（Artificial Intelligence, AI）凭借对多源异构数据的

收稿日期：2026-02-09

基金项目：国家自然科学基金项目（32300248）；山东省自然科学基金项目（ZR2022QC010）

作者简介：徐启玉，博士，副研究员，研究方向为作物分子育种。E-mail：xuqiyu@nwafu.edu.cn

*通信作者：郑 波，硕士，高级教师/副研究员，研究方向为农业科技与作物育种。E-mail：bo.zheng@pku-iaas.edu.cn

copyright©2026 by the authors （登录 www.smartag.net.cn 免费获取电子版全文）

深度特征提取与建模能力[10, 11]，为突破传统育种瓶颈提供了一条变革性技术途径，正成为连接大数据与育种决策的关键桥梁[12]。

在当前智慧育种体系中，AI驱动的算法矩阵正呈现出清晰的层次化技术架构与垂直化演进态势。作为AI这一广义技术范式中专注于多层神经网络构建的核心子集，深度学习（Deep Learning, DL）凭借极强的非线性特征提取能力，已成为全基因组预测和表型解析的重要引擎[13]。随着大参数通用化预训练范式的兴起，DL逐步跃升为具备跨任务泛化能力的基础模型。其中，擅长处理序列语义与文本逻辑的大语言模型（Large Language Models, LLMs）构成了基础模型最具代表性的技术分支。而当这种基于大规模语言建模的技术内核向种业垂直领域深度渗透，并与作物基因型、表型组、环境组等多模态大数据深度融合时，便催生了专门解析复杂数量性状的“育种大模型”。这一由底层通用算法向智慧种业垂直应用延伸的层级架构，赋予了现代育种系统强大的数据洞察与逻辑推理等涌现能力[14]。在实际育种场景中，这种协同应用的核心价值在于：一方面，通过整合气象、土壤及酸碱度等环境组学信息，实现特定生境下的品种精准推荐与栽培方案优化；另一方面，利用计算机视觉技术实现产量相关表型的大规模量化与自动化误差矫正；最重要的是，AI能基于亲本间的基因型差异与表型互补性，实现杂交组合表现的反向推测与虚拟筛选。这种转变标志着育种范式已从经验驱动跨越至智能设计新阶段[14, 15]。

本文旨在系统探究AI驱动植物育种范式变革的底层逻辑。文章首先梳理了从经验育种到智慧育种的代际演变，重点阐述AI在表型组学、基因组预测及分子设计等关键环节的技术进展；随后，深入探讨以“育种大模型”为核心的智能系统如何重构育种全流程；最后，针对当前领域存在的数据壁垒、模型可解释性等瓶颈进行讨论，并展望未来发展方向，以期为中国种业科技实现自立自强提供参考。

1 植物育种的演变历程

植物育种是伴随人类文明最古老也是最重要的活动之一，经历了从自然到科学的显著演变。植物育种的历史演变如图1所示，通常可以分为五个代际[16]。

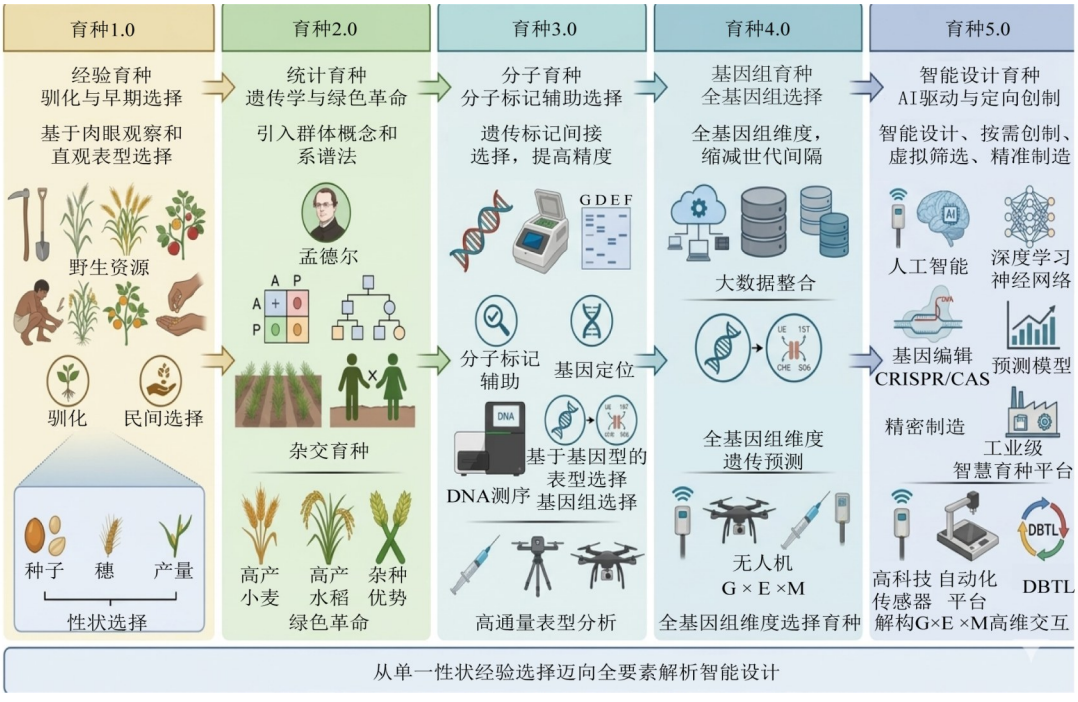


图1 作物育种技术的历史发展

Fig. 1 Historical perspective on crop breeding technology

1.1 育种1.0:作物驯化与经验育种

该阶段始于约一万年前的作物驯化[17, 18]。由于缺乏系统理论与方法，人类只能根据肉眼观察和经验积累，选择落粒性低、果实大、生育期短等直观

表型，再经长期人工驯化获得性状优良的品种[19]。这一阶段本质上是对自然变异的偶然捕捉，遗传增益的积累极其缓慢且不可控[20]。

1.2 育种2.0：常规育种与绿色革命

19世纪末孟德尔遗传定律的发现，标志着育种进入2.0阶段[21]。育种家开始利用有性杂交整合优良等位基因，配合系谱法等选育方法，使目标性状的选择具备了初步的可预测性[22]。20世纪40—60年代，以半矮秆基因大规模利用为核心的“绿色革命”，通过降低株高增强抗倒伏能力，以及化肥和农药的应用，显著提升了全球作物的单产潜力，解决了基础性的粮食短缺问题[23, 24]。此后，诱变育种技术的兴起进一步拓宽了遗传多样性，打破了自然变异的局限[25, 26]。

1.3 育种3.0：分子标记辅助与基因工程

20世纪末，随着分子生物学和生物技术的飞速发展，育种进入了分子标记辅助选择与基因工程驱动阶段。1977年第一代测序技术的问世为精准育种奠定了基础。相比传统育种，该阶段实现了从个体表型选择向单基因/主效数量性状基因座（Quantitative Trait Loci, QTL）选择的跨越[27]。然而，该阶段在处理受微效多基因调控的复杂性状时仍面临瓶颈，且难以解析复杂的G×E交互效应[28]。

1.4 育种4.0：分子设计育种与全基因组选择

21世纪初，全基因组选择（Genomic Selection, GS）概念的提出及高通量测序成本的降低，推动育种进入4.0阶段[29-31]。2001年GS概念的提出及后续高通量测序技术的降本增效，使育种首次实现了全基因组维度的遗传预测，大幅缩减了世代间隔。然而，随着研究对象从单一基因扩展到转录组、代谢组及表型组，育种数据呈现出海量、高维及非线性的特征，传统的线性统计模型在处理非线性关联和复杂交互效应时遭遇了瓶颈。

1.5 育种5.0：AI驱动的智能设计育种

目前，育种科学正向以AI为核心驱动力的5.0阶段跨越，亦被称为“智能设计育种”或“5G育种”，其核心特征是生物技术（Biotechnology, BT）、信息技术（Information Technology, IT）和高通量组学大数据的深度融合。作为一种模拟、延伸和扩展人类智能的新技术科学，AI在5.0阶段扮演了“决策大脑”的角色[11]。一方面，通过集成环境组学的预测模型（Integrated Genomic-Enviromic Prediction, iGEP），AI能够把气象、土壤、遥感等多维环境因子整合起来，破解复杂性状在不同生态区下的动态表达，实现品种的精准推荐与环境适配[32]。另一方面，LLMs与CRISPR/Cas9等精准编辑技术的结合，使育种从大海捞针式的筛选转变为按需设计的精密制造[33, 34]。5.0育种不仅优化了杂交育种的底层组配逻辑，更利用AI的预测能力在虚拟空间内对数以万计的亲本组合进行预筛选。随着工业级智慧育种平台的出现（如Smart Breeding Platform）[35]，人类正跨入精准预测、虚拟筛选与性状定向创制并行的全新纪元。

2 AI赋能作物育种的研究进展

AI在作物育种领域的研究已由早期的单一算法辅助分析，发展为涵盖种质资源评价、高通量表型解析、遗传基础挖掘和育种决策的全流程智能体系[36, 37]。如图2所示，该流程遵循“数据驱动-模型理解-决策反馈”的逻辑，底层为数据感知与多模态表征：整合了覆盖作物全生命周期的多组学数据、高通量表型数据及动态环境数据，通过多源异构数据的标准化表征，为后续模型提供多维度的信息输入。中层为认知引擎与智能建模：利用DL、LLMs及多模态融合技术，实现从海量原始数据到深层生物学语义的解析，捕捉复杂的基因型-表型关联。顶层为决策反馈与性状创制：将中层挖掘出的规律转化为种质资源高效利用、性状精准预测及基因定向设计等具体育种行为，形成智慧育种的全景蓝图。

相较于依赖显性假设与线性回归的传统生物统计学，以DL和LLMs为核心的AI提供了一种全新的数据驱动方法论，通过在海量异构数据上进行无监督或自监督训练，构建了从底层资源挖掘到高层智能设计的逻辑闭环，极大提升了育种的精准化程度（图2）。

2.1 种质资源挖掘与评价

种质资源是育种创新的物质基础。AI正在深刻变革大规模基因组数据的表征方式与组织逻辑，推动种质评价向动态演化预测转变[38]。利用DL的非线性建模与特征压缩能力，育种家可对全基因组范围内的变异进行高维模式识别[13, 39]。其核心突破在于：借助自监督学习算法，AI能够从海量未鉴定的序列中提取隐含的生物学特征，挖掘出传统统计学难以察觉的罕见等位基因与优异变异。这不仅提升

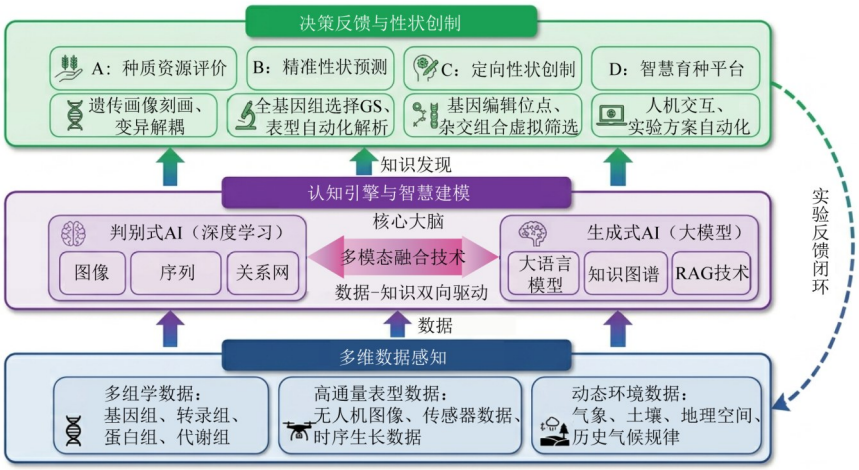


图2 AI赋能作物育种路线图

Fig. 2 Roadmap of AI-enabled crop breeding

了泛基因组构建与种群结构解析的精细度，更有效激活了种质库中的潜在遗传价值。

种质评价不仅依赖于田间实测，还依赖于对其历史表现和遗传背景的深度理解。LLMs的跨界应用为挖掘散落在海量文献、历史档案中的隐性评价信息提供了变革性手段[12, 14, 40]。比如，采用Transformer架构的PlantConnectome等模型，能自动从数以万计的学术文献中提取出特定种质的基因-性状关联，构建起结构化的知识图谱[41]。这使得原本孤立的种子信息转化为具备生物学意义的评价数据，实现了对种质资源历史表现的自动化评价。

此外，针对特定作物的垂直领域研究也取得了显著进展。例如，小麦种质信息提取（Wheat Germplasm Information Extraction, WGIE）等对话式大模型，通过检索增强生成（Retrieval-augmented Generation, RAG）技术，从零散的育种记录中直接提取出抗病性、丰产性等关键背景信息[42]，有效缓解了数据碎片化问题，为育种家提供了具备逻辑推理能力的种质资源评价参考。

传统的种质评价往往是静态的，难以预判资源在不同生态区或未来气候下的表现。AI通过引入环境演化维度，实现了种质资源从静态描述向动态评价的跨越[43]。通过整合历史气象、土壤理化指标及种质演化轨迹，AI模型可以模拟特定种质在极端气候或不同种植场景下的表现。这种基于G×E互作的预测性评价，为种质资源的异地引种和精准选配提供了前瞻性的科学依据。

综上，AI的应用使种质资源库由原来的种子库变成了智能化评价系统。通过为每份资源精准刻画“遗传画像”，AI实现了对复杂变异与环境交互效应的深度解耦。这不仅降低了获取种质核心价值信息的门槛，也为后续的精准育种奠定了坚实的评价底座。

2.2 高通量表型精准解析

表型鉴定是现代育种进程中限制遗传增益提升的主要瓶颈之一[44]。随着计算机视觉和DL技术的深入应用，表型分析正由依赖人工感官的离散评价，转变为高频率、非破坏性、多维度的全自动解析过程[45]。如图3所示高通量表型解析的标准化技术工作流程，该过程利用工业级传感器捕捉作物生命全周期的多源信号，通过算法将其转化为定量的生物学指标，从而在数字空间内精确重构“基因型-环境-表型”的动态交互关系。

高通量表型解析的首要环节是创建高保真的物理实体镜像，LemnaTec、PhenoSpex等工业级平台改变了传统依靠人工尺量或者目测的作业方式（图3）[46]。以PhenoSpex平台的PlantEye F600多光谱3D成像技术为例，其利用激光扫描与微秒级反射信号获取技术，可生成空间分辨率达0.1 mm的高精度3D点云模型。这不仅能实现株高、叶面积、倾角等形态学参数的快速提取，还能通过同步获取的近红外和红光波段数据，实时计算归一化植被指数（Normalized Difference Vegetation Index, NDVI）以评估植株生理状态[47, 48]。在此基础上，数字孪生技术的引入使育种家能在虚拟空间内模拟作物在不同生长阶段对养分、水分的需求，为全生育期动态表型的精准解析奠定标准化数据基础[49]。

在数据感知与表型重构层面，卷积神经网络（Convolutional Neural Network, CNN）及其变体，

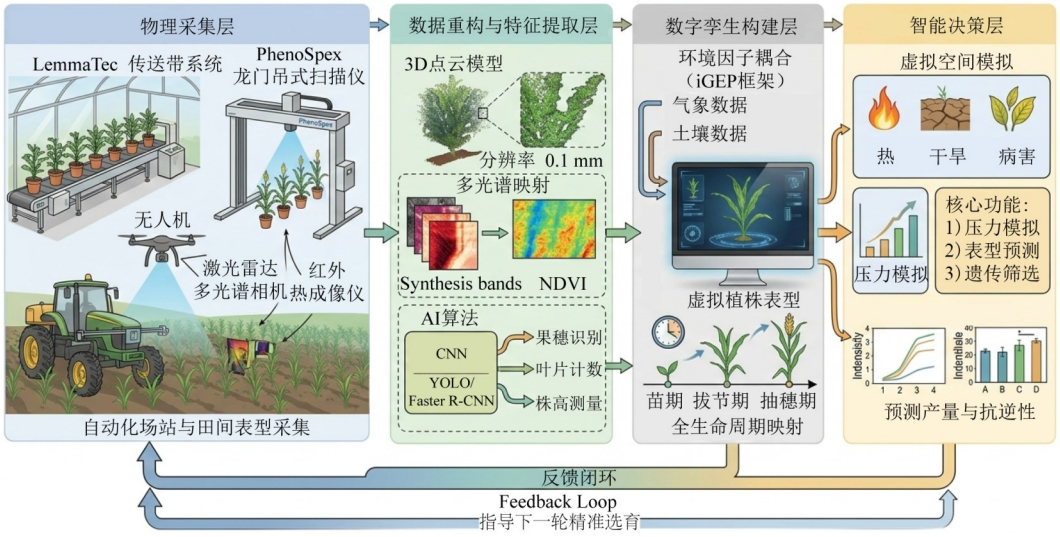


图3 作物高通量表型解析与数字孪生工作流程

Fig. 3 Crop high-throughput phenotyping and digital twin workflow

如更快的基于区域的卷积神经网络（Faster Region-based Convolutional Neural Network, Faster R-CN；或单阶段目标检测模型（You Only Look Once, YOLO）等，已不仅是解析农田无人机影像和传感器成像数据的主要工具[50-52]，更是构建作物数字孪生（Digital Twin, DT）实体的核心技术桥梁（图3）。在水稻、小麦等作物育种中，DL模型通过自动识别并统计果穗数量等形态特征，实现了对作物物理形态的数字化映射[53]。

进一步地，如图3所示的多模态特征耦合，AI能够解析人类视觉难以察觉的生理反馈，赋予DT实体生理感知能力。例如，通过热红外图像捕捉根系吸水受阻引发的叶片温度异常[53]，并通过集成iGEP框架，将气象、土壤等环境因子与图像特征深度融合。这种跨模态的解析能力，使得育种家能够在虚拟空间中构建全生命周期的作物动态DT模型，从而识别出作物在特定逆境下的遗传潜力[32]。这种跨模态的解析能力，为耐逆性品种的选择提供精准的定量依据。

最终，AI将海量原始信号转化为具备生物学意义的定量指标，如生物量预测、抗病性分级及环境适应性评价（图3）。针对育种数据存在的样本匮乏与标注不一问题，生成对抗网络（Generative Adversarial Network, GAN）和LLMs发挥了关键作用。GAN通过模拟生成不同环境压力下的作物生长图像，有效扩充了极端逆境下的稀缺训练样本，显著提高了模型的泛化能力[54]。LLMs凭借其出色的语义理解能力，能够将不同地区、不同育种家持有的主观差别观察笔记自动映射到统一的表型本体中[28, 40]。

综上所述，图3所示的技术流通过语义对齐与数字化表征，实现了大规模跨环境下的G×E交互分析。这不仅实现了历史数据的再利用，更通过高质量的数据输入为后续的全基因组选择模型提供了有力支持[55]。

2.3 基因组数据解释与预测模型演进

在GS领域，AI正在促使模型由传统的线性统计向深度非线性建模转变。这一转变不仅提升了预测精度，更借助于AI强大的模式识别能力，实现了对高维遗传交互作用与复杂调控逻辑的深度解释[56-58]。

传统GS模型如基因组最佳线性无偏预测（Genomic Best Linear Unbiased Prediction, GBLUP）建立在加性遗传效应假设之上，将复杂的上位性与显性效应视为不可解释的残差[59]。其数学表达如公式（1）所示：

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \quad (1)$$

式中：$\boldsymbol{y}$为观测值向量；$\boldsymbol{b}$为固定效应向量；$\boldsymbol{u}$为加性遗传效应向量；$\boldsymbol{X}$为固定效应设计矩阵；$\boldsymbol{Z}$为随机效应设计矩阵；$\boldsymbol{e}$为残差向量。在这样的线性范式之下，模型假定所有的单核苷酸多态性（Single Nucleotide Polymorphism, SNP）位点效应都服从正态分布，主要处理基因间的加性遗传效应。但是作物重要的农艺性状常常受到复杂的上位性和显性效应调控，非线性交互作用在传统线性模型里一般被当作无法解释的残差，限制了模型对复杂关联的表示能力。AI技术的介入填补了这一解释空

白。以多层感知器、随机森林、梯度提升机等为代表的算法，利用非线性激活函数和决策树分裂机制，能够自动捕捉基因间的上位性互作。这种建模方式的进步，使得研究者能够解释被线性模型忽略的非加性效应，从而辅助育种家从成千上万个候选亲本中筛选出最佳组合，解决传统组配实验规模过大的难题。

DL架构进一步实现了对基因组拓扑结构的生物学解析。例如，最近开发的大豆基因组预测深度网络框架（Soybean Deep Network for Genomic Prediction, SoyDNGP）通过构建3D CNN，将SNP位点按照其在染色体上的物理空间关系进行编码，打破了传统模型将SNP视为独立、无序变量的局限，实现了对大豆复杂性状的高精度预测[60]。通过卷积核的特征提取，模型能够识别出具有功能相关性的基因组邻域，从而解释了连锁不平衡单倍型与复杂性状之间的空间关联逻辑。

针对跨生态区育种中的G×E交互难题，Transformer架构及其变体，如深度表型关联特征提取模型（Deep Pheno Correlation former, DPCformer）的引入提供了全新的语义解释维度[61]。Transformer的核心贡献在于其自注意力机制，它能动态计算不同遗传位点在特定环境背景下的贡献权重[61]。通过整合特定地区的历史气象（温光水）、土壤营养状态及酸碱度（pH）等环境特征向量，Transformer模型能够实现对特定生境下品种表现的精准预测与推荐。通过可视化注意力地图，育种家可以直观地观察模型在不同环境压力下（如干旱或高温）分别关注了哪些关键遗传区段，为解释复杂性状的环境响应机制并进行栽培方式的预测优化提供了强有力的工具。

然而，对于百万级碱基对长度的原始基因组序列，Transformer架构存在严峻的计算瓶颈。由于其核心自注意力机制的计算复杂度和序列长度N呈二次方关系O（N^2），直接处理完整的染色体级别数据在算力资源上具有很高的挑战。为了克服上述限制，研究者通过结合Transformer的思想和卷积的计算效率，创造了基于Hyena Operator的架构，并将其应用在基因组学领域，得到了HyenaDNA预训练大模型[62]，使得AI模型具备了解析染色体级别超长序列的能力。该技术逻辑重构在于用长卷积滤波器代替传统的注意力矩阵计算，用频域快速傅里叶变换处理把计算复杂度从原来的O（N^2）降到O（$N\cdot\log N$）。效率质变之后，该模型不再仅关注单个SNP，而是能够解析跨越数万碱基的启动子、增强子及远端调控元件之间的协同逻辑[63]。这种对深层生物学语义的获取，使得AI能够像理解自然语言语法一样，解释基因组的编码规则。基于此，AI可以根据不同亲本间的基因型差异，反向推测出实现目标产量所需的最佳等位基因组合，实现从“目标表型”到“理想亲本配对”的反向选配设计[64, 65]。

综上所述，AI在基因组领域的应用已超越了单纯的数值计算，正演变为一种探索生命复杂系统的方法论。通过CropGS-Hub、Smart Breeding Platform等平台的集成化应用，AI实现了预测精度与生物学解释的双重驱动，通过对环境条件预测、表型误差校准以及最佳组合推荐的深度整合，为解析复杂农艺性状的育种决策提供了全新视角[35, 66]。

2.4 多组学数据整合与系统生物学推断

针对作物复杂性状背后“基因-转录-蛋白-代谢”的多层级调控，AI模型的整合策略直接决定了预测的稳健性与生物学解释深度。目前，主流的多组学融合架构根据其干预阶段的不同，演化出三种逻辑路径[67, 68]。

早期融合：将不同维度的组学特征在输入层直接拼接。这种方式虽然保留了原始数据颗粒度，但由于离散的SNP序列和连续的代谢组、环境组数据在量纲和稀疏性上差异巨大，模型易出现特征覆盖现象，导致泛化能力下降。

晚期融合：采用集成学习逻辑，各组学数据分别训练专门的子模型并在决策层加权。该结构虽然具有很高的并行度，但由于各层级间被物理隔离，模型难以捕捉跨组学的非线性调控效应[69, 70]。

中期融合：这是目前育种模型的主流演进方向。其核心是在深度神经网络的隐藏层中实现特征投影和交叉，构建一个共同的潜在表示空间，使模型可以学习基因、转录、代谢之间的动态协同关系，能够有效捕捉到跨层级的生物学补偿效应[71]。这种架构的优越性在植物全基因组预测模型（Deep Neural Network for Genomic Prediction, DNNGP）中得到了充分体现[72]。DNNGP利用多层神经元构建复杂的非线性投影，有效捕捉到了传统模型难以刻画的跨层级的生物学补偿效应。DNNGP利用联合嵌入技术，将离散的SNP序列与连续的代谢组、环境组特征映射到同一高维语义空间。这种处理方式突破了不同信息载体之间的量纲

限制，实现了对复杂性状的深层次语义建模。在这种范式下，AI不再仅仅是对抽象的数字矩阵进行统计相关性分析，而是通过捕捉基因型与代谢组、环境组之间的内在紧密联系，实现了从单一组学预测向全系统智能推断的跨越。这极大增强了机器对生物复杂系统的理解能力，使其能够模拟生物体在面对遗传或环境扰动时的系统性响应[72]。

多组学整合面临的最大技术难题在于如何将数学维度的权重分布转化为具有生物学意义的可利用信息。针对传统黑盒模型难以解释性状形成机制的瓶颈，“大语言模型+知识图谱”的深度耦合成为核心解决方案。在这种架构之下，LLMs作为“一体化中枢”，凭借强大的文本理解与逻辑推理能力，将各组学的观测值进行语义化建模，把繁杂的调控网络转变为可理解的科学假说。其次，图神经网络被用来在结构化的知识图谱上进行多跳推理[73, 74]。该链路首先是将AI预测的候选基因锚定在知识图谱节点上，随后搜索其通过转录、翻译和代谢反应通向目标性状的逻辑路径，最后结合实验数据完成因果发现[75]。通过这种深度耦合，AI模型能够跨越传统统计学的关联限制，挖掘出复杂生物网络中最核心的因果变异。进一步结合作物生长模型，AI可以模拟不同育种决策在特定环境下的遗传增益潜力[75]。这不仅为亲本选配提供了确定性的数据支撑，更通过整合多维生物学知识库，促使现代育种朝着全生命周期的系统设计时代迈进。

2.5 分子设计育种

在分子设计育种范式下，解析复杂的“序列-功能”映射图谱是实现按需育种的主要难题之一[40]。AI技术正通过对基因组调控元件的深度解构，针对性地解决作物育种中的关键科学问题，如产量潜力的精准预测、特定生态区的品种推荐以及海量杂交组合的辅助决策，驱动育种从随机筛选向确定性的智能设计转变。

识别顺式调控元件与基因表达之间的逻辑联系是设计育种的基础，也是解决产量相关表型解析的关键。目前，以先正达集团与AI公司InstaDeep合作开发的农业核苷酸变换器大模型（Agro-Nucleotide Transformer, AgroNT）为代表的植物语言大模型，利用DL技术对跨物种的非编码序列进行预训练，实现了对启动子强度、组织特异性表达模式的零样本预测[76]。该技术为育种过程中成千上万的潜在调控位点提供了科学的优先级排序，通过整合特定地区的历史气象、土壤氮（N）、磷（P）、钾（K）养分及pH等信息，AI能够预测调控元件在特定环境下的响应强度，从而实现对特定品种在特定生境下表现的精准评价与栽培方式的预测优化。此外，跨物种迁移学习揭示了生命语言的共性规律。例如，FloraBERT模型通过共享不同植物间的序列语法信息，成功将模式植物中习得的调控逻辑迁移至遗传背景更为复杂的作物中[77]。这种迁移学习策略使得AI能够在不同作物的遗传底盘上完成调控元件的功能重组，极大提升了对未知基因功能的预测效率，为解析不同亲本间的表型差异提供了分子证据[77]。

当前，AI正在促使基因编辑技术由工具筛选向设计工具迈进[76, 78]。例如，以OpenCRISPR-1为代表的大模型实现了逻辑上的重大突破：它是全球首个完全由AI大语言模型从头设计的基因编辑器。其逻辑内核在于通过大规模语言建模学习蛋白质和DNA交互的底层物理化学语法，而非仅仅依赖自然界的进化产物[79]。由于该模型具有强大的涌现能力，AI可以直接生成自然界中不存在但具有高潜在活性的编辑蛋白[79]。在育种实践中，这种能力可直接用于解决产量形成的科学瓶颈：针对性地设计具备高潜在活性的编辑蛋白，对影响产量的关键基因（如穗发育相关基因）进行精准改良，从而绕过传统杂交育种中优异等位基因聚合困难的问题，显著提高变异创制的特异性[33, 34, 79]。

针对育种家最为关心的“最佳组合推荐”难题，AI利用高维非线性建模能力，正在重新定义杂交育种的筛选逻辑。传统育种每年需进行几万个组合实验，耗费巨大。AI通过将异构组学数据进行语义化整合，并执行基于知识图谱的多跳推理，实现了海量基因型组合的虚拟筛选[12, 40]。AI不仅能根据亲本间的基因型差异进行配合力评估，还能根据目标产量性状进行“反向推测”：即根据预设的目标表型，反向检索最佳亲本配对，实现在计算机内部对万级组合的快速筛选与精准评价。这种架构将预测结果同已有的生物学通路联系起来，通过对表型数据的自动化校正与量化，从复杂的网络中找出影响农艺性状的深层因果变异和关键调控因子[13]。

综上所述，AI创建了一个高度集约的虚拟育种闭环。它不仅为创建智慧育种提供了理论依据，更通过整合计算预测与生物创制，将育种从大海捞针式的经验摸索转变为确定性的智能制造流程，通过优化选配逻辑大幅缩减了育种周期与投入[34]。

3 作物育种大模型发展现状

3.1 育种基础模型的涌现逻辑与参数扩展

作物育种大模型是指基于AI技术，特别是深度学习与大规模数据训练，专门用于支持农作物育种全过程的智能系统[80, 81]。这类模型的出现与兴起意味着农业AI的全面技术飞跃。

育种大模型普遍采用Transformer架构，如前所述，该架构的核心逻辑在于自注意力机制[82]。该机制通过动态计算序列内部元素间的相互关注度权重，赋予了模型解析高维组学数据的卓越能力。在育种场景中，该类模型能够有效捕捉跨染色体的基因调控逻辑以及复杂的遗传互作关系（如上位性效应），实现了从文本语义到生物学功能语义的深度逻辑迁移[83]。

大模型在育种领域中的效能提升遵循严格的扩展定律，这决定了智慧育种算力资源投入的科学边际。研究表明，神经网络性能（L）与其参数规模（P）、数据集规模（D）、训练计算量（C）之间存在幂律关系，即KM法则[84]和Chinchilla法则[85]。其典型性能如公式（2）～（4）所示：

$$L(P)=\left(\frac{\mathrm{P_c}}{P}\right)^{a_P},\ a_{P\sim0.076} \tag{2}$$

$$L(D)=\left(\frac{\mathrm{D_c}}{D}\right)^{a_D},\ a_{D\sim0.095} \tag{3}$$

$$L(C)=\left(\frac{\mathrm{C_c}}{C}\right)^{a_C},\ a_{C\sim0.050} \tag{4}$$

式中：$\mathrm{P_c}$、$\mathrm{D_c}$、$\mathrm{C_c}$为常数，分别对应不可约损失的临界截断值；α_P、α_D、α_C分别为参数、数据和计算量的扩展指数，决定了性能随规模扩张的边际递减速度。

这就意味着在既定的算力预算下，模型规模和数据量必须等比例扩张，才能获得最优的遗传增益预测效果[85]。在育种实践中，构建大规模参数模型是解析基因型、表型，以及环境等诸多高复杂度交互系统的必然选择。当模型规模突破临界阈值时，会产生小规模模型所不具备的涌现能力，包括上下文学习、指令遵从、逐步推理等[86]。在智慧育种场景中，大模型不仅能通过记忆已知遗传规律，更能通过思维链技术处理未知的复杂推理任务。例如，在预测多基因编辑组合对植株综合性状的非线性影响时，模型能自主生成中间推理步骤，给出具备生物学逻辑支撑的育种方案。这种模式显著超越了传统DL模型仅依赖局部模式识别的局限[87]。

支撑该复杂能力的基础是自监督学习技术，它直接从没有标注的海量基因组、表型组、气象环境数据中学习生命的底层逻辑，有效解决了农业领域高质量标注数据稀缺的痛点。结合3D并行技术与零冗余优化器（ZeroRedundancy Optimizer, ZeRO），育种大模型能够在分布式GPU集群上实现高效参数扩展，将零散的多组学数据转化为可推理、可设计的认知引擎[88]。此外，这种工业化底座也为模型从可解释向可信任的治理架构转型提供可能，确保AI生成的育种决策在实际田间环境下既稳健又安全。

3.2 语言大模型与农业知识体系重构

在现代育种研究中，LLMs正演变为将海量非结构化文献转化为结构化知识发现的认知中枢。其核心意义不仅在于对词序列的概率建模，更在于对生物学语义的深层理解与逻辑推演，从而重塑传统碎片化、静态化的农业知识结构[82]。

从技术架构上来看，农业领域大模型的发展经历了从双向编码器架构向因果解码器架构的转变。早期代表性如农业双向编码器表征模型（Agricultural Bidirectional Encoder Representations from Transformers, AgriBERT），利用BERT结构对海量农学语料进行预训练，并通过农业专业知识图谱进行微调。其侧重于命名实体识别与文本分类，初步提升了农业要素的匹配精度[89]。随着参数规模的提升，以大语言模型Meta注意力机制2代（Large Language Model Meta Attention 2, Llama 2）为底座的PLLaMa系列因果解码器架构模型展现出卓越的少样本学习和泛化能力[90]，渐渐成为农业垂直大模型的主流选择。国内研发的“司农”大模型通过挖掘大量的农业专著、学术论文，试图解决通用模型在农业上出现的幻觉问题；“神农”大模型则是进一步将知识图谱和向量数据库结合起来，大大提高了模型的可靠性。这些模型演进的本质是将模型从简单的信息检索器升级为逻辑推断器，为解决复杂的育种因果关系提供了底层算力支持。

为了克服通用模型在农业垂直领域的“幻觉”问题，RAG技术为模型提供了动态的知识获取路径[91]。RAG系统通过将农业问题转化为高维空间的向量表征，利用基于嵌入的检索在权威文本数据库中精准定位相关片段，并将其作为背景提示输入模型。这种机制确保了LLMs在确定的语义范围内

进行回答，配合思维链等分步推理技术，模型能够将非结构化的育种记录自动转化为结构化的实体关系网络[87]。这种知识锚定逻辑不仅提高了模型在具体育种场景中的应用精度，更将可解释性嵌入生成路径中，使复杂的育种建议具备了可追溯的证据链。

通过LLMs的深度介入，传统的静态知识库就转变成了具有语义联系、能够进行复杂逻辑推理的动态认知引擎[40]。在综合化农业应用场景中，LLMs不仅是智能问答的核心，更是育种决策的中枢。它可以通过接入物联网实时监测气象、土壤及设备状态，并凭借强大的指令执行能力为育种家提供个性化的科学建议。在这一动态重构过程中，通过整合高质量数据体系与版权保护机制，大模型确保了育种知识的合法流通与高效利用。这种深层次的知识重构极大地减少了对人工经验的依赖，促进了农业知识的指数积累与精准传播，为实现全流程、智能化的智慧育种服务奠定了坚实的技术基础。

3.3 多模态融合与育种决策智能体的演进

随着现代育种从单一信息载体的智能设计向多模态大模型的跨越式发展，语言、图像及结构化组学数据之间的语义对齐已成为破解农业信息孤岛的关键。不同于仅能处理单一类型数据的孤立模型，多模态架构旨在通过统一的嵌入空间将异构信息进行深度融合，从而显著提升机器对复杂生物系统的理解与建模精度。跨媒介的知识融合不仅是通用AI在农业领域落地的必要门槛，更是构建精准育种决策系统的主要驱动力。其技术本质在于利用预训练的视觉编码器和语言编码器，将高通量表型图像、育种文献、历史记录及微观组学数据映射到共同的潜在表示空间，进而学习不同模态间的关联规则与生物学映射关系[92]。在此架构支撑下，模型能够实现以文寻图或以图释理的跨模态推断，构建起集成基因组-环境组预测的宏大框架，使其在解析微观遗传变异的同时，能够敏锐捕捉宏观物候特征与时空多变的环境压力。

随着模型逻辑推理能力的提高，育种决策系统正由被动响应的工具演进为具备自主规划能力的育种智能体。作为智慧育种的大脑，育种智能体能够实时接入物联网监控设备、气象预警及其他专业人工智能模型，将多源异构的输入信号转化为实时最优的育种策略或者栽培方案。与传统的规则驱动系统不同，智能体具备极强的指令执行与逻辑推演能力，不仅能通过多轮对话将深奥的遗传学理论转化为个性化的行动建议，还能够在虚拟实验室中进行大规模杂交组合的模拟预演。这种演进标志着AI在农业领域的应用已从局部的模式识别阶段，跨越到了系统化的决策支撑与方案规划阶段，极大地压缩了育种研发成本。

由崖州湾国家实验室、中国农业大学、上海人工智能实验室联合开发的丰登（SeedLLM）便是这一演进路径的典型代表[40]。该模型于2024年4月发布，于2025年7月完成智能化升级。作为农业AI领域的里程碑，“丰登”已从最初的专业知识问答模型，进化为集品种选育、种子经营、栽培管理于一身的综合性AI智能体。它不仅能够应对复杂的遗传学理论问答，更具备完成亲本选择和杂交设计的自主决策能力。其核心优势在于对海量农业多组学数据进行了高效索引和语义重构，通过将预测模型与专家知识库深度耦合，为科研人员和农户提供准确的品种推荐和种植指导[66]，为构建全球粮食安全保障体系提供了强有力的技术引擎。

3.4 垂直领域模型的适配与轻量化路径

由于农业生态的环境高度异质性与作物种类的多样性，通用大模型往往难以直接涵盖所有细分育种场景下的时空变量。因此，构建具备地域特色或者物种特性的垂直领域模型成为必然选择。然而，大模型的预训练和全参数微调极度依赖高性能GPU服务器和计算资源，这为育种机构的实际部署带来了成本与技术的双重挑战[93]。研究高效的模型适配技术和轻量化途径，不仅是缓解算力瓶颈的工程手段，更是驱动AI育种由实验室走向产业化的关键。

为了在不改变模型核心参数的前提下使其快速具备领域专业能力，参数高效微调（Parameter-Efficient Fine-Tuning, PEFT）技术得到了广泛应用。相比计算开销巨大的全参数微调，PEFT仅需调整少数额外参数即可实现下游任务的适配[93]。其中，适配器微调通过在注意力层和前馈层之间插入小型神经网络模块，显著提升了模型对作物病虫害识别等特定任务的灵敏度。除此之外，前缀微调和提示微调则分别用固定向量或者提示模板来激发模型在育种任务上的性能。目前应用较为广泛的低秩适应（Low-Rank Adaptation, LoRA）技术，则通过在Transformer层中引入低秩分解矩阵来模拟模型内部

较低的本征维度，从而以较小的参数扰动实现高效学习[94, 95]。这种微调范式不仅提高了模型对特定作物生物学规律的对齐精度，更增强了其在罕见等位基因预测、极端逆境响应等复杂场景下的泛化稳健性，为解析复杂的基因型-环境互作提供了低成本且高可靠度的路径。

在解决算力瓶颈方面，配套的并行化训练与模型压缩技术构建了智慧育种的工业化底座。通过集成数据并行、流水线并行与张量并行的3D并行技术，配合ZeRO技术，系统能够有效消除多GPU协作中的数据冗余并大幅降低显存需求[88]。这种架构优化使育种平台能够实时处理海量多组学数据流，并在“设计-构建-测试-学习”（Design-Build-Test-Learn, DBTL）循环中实现快速迭代。在模型压缩方面，LORA等量化技术将参数精度从FP16压缩至INT4，在保持模型预测性能基本稳定的前提下，极大地缩减了存储体积[96]，结合混合精度训练等手段，数据传输与计算效率显著提升。

目前垂直领域的农业大模型正朝着轻量化、便携化方向快速发展。以2026年初发布的“司农”大模型8B、32B版本为例，其核心目标在于降低中小型种业企业与科研院所的准入门槛。这种部署模式配合国家级数据版权一体化服务平台的构建，有效解决了育种数据要素流转中的合规性与安全性问题。未来，随着算法与算力硬件的协同进化，轻量化育种大模型将越来越多地嵌入手持设备或无人农机终端，与农田感知物联网实时对接。这种由云计算向端侧感知的转变[97]，将赋予育种家实时、个性化的精准决策能力，甚至引领农业领域向无人化精准操作发生变革性飞跃。

3.5 育种能力基准评估与可靠性保障

在利用AI驱动育种决策的过程中，幻觉现象，即模型生成看似逻辑自洽但违背生物学事实的信息，已成为制约其在科研和生产领域大规模应用的底层红线。由于作物育种具有高投入、长周期及环境高度敏感等特性，任何错误的遗传学建议都可能导致田间试验失败与珍贵种质资源的不可逆浪费。因此，构建标准化的育种能力评估基准和多层级可靠性保障机制，是确保AI真正成为育种专家的必经之路。

为了系统评价育种大模型处理复杂专业问题的能力，研究人员提出了如SeedBench等多任务评估基准[98]。该基准涵盖了从基因检索、表型解析到品种选育、方案设计及农艺特性优化的全流程测试任务，深度考察模型在功能调控机制推理方面的表现。此类测评框架的主要意义在于通过量化反馈识别模型的知识缺陷与潜在偏见，从而指导模型朝着符合生物学逻辑的方向迭代[98]。在此过程中，评价维度正从传统的文本匹配准确度转向生物学因果发现能力，重点考察模型在处理基因型、环境等复杂交互作用时是否具备真实可靠的涌现能力。

实现可靠性保障的首要技术路径就是将大模型的语义生成能力与确定性的外部知识库进行深度耦合，这一过程主要依靠RAG技术实现[91]。通过构建包含专利、学术文献及育种档案的高质量向量数据库，模型在生成建议前会预先进行语义检索，将权威数据作为背景约束输入。这明显改善了模型对特定育种问题的回答准确率，有效规避了生成过度泛化或误导性建议的风险[91]。此外，可靠性不但取决于算法改进，更依赖于高质量AI数据体系的支撑。通过构建国家级数据版权一体化服务平台，确保输入RAG系统的每一条育种数据均具备来源可追溯性与法律合规性，能够从源头上消解因数据污染带来的技术风险[99]。

为了确保输出结果符合育种专家的专业预期，还需引入多维评估机制以及对齐技术。在评估层面，利用GPT-4等高级模型配合人工标注，从逻辑流畅性、生物学合理性及因果一致性等维度对生成方案进行闭环评价[100]。在优化层面，则采用对齐微调的方法，例如基于人类反馈的强化学习[101]，将育种家的科学经验、价值偏好植入模型底层逻辑，严防模型生成带有生物安全性风险或遗传偏见的内容。

综上，育种大模型的可靠性保障是一项集成数据锚定、逻辑推理与专家经验对齐的系统工程[102, 103]。通过构建可信任、可解释且稳健的AI系统，育种家能够在显著降低实验成本的同时，获得具有深刻生物学洞察力的决策支持。这种由技术解释转向多维信任的治理改变，正推动现代育种从经验驱动向智能设计驱动发生本质跃迁，为构建智慧育种新生态提供坚实的理论支撑。

4 AI驱动育种:模型演进、实证应用与效能评价

全球作物育种正处于数智化竞争的关键窗口期，大模型技术凭借其在复杂关联表示与多模态信息处理上的显著优势，已成为种业强国博弈的核

心。以拜耳、科迪华为代表的国际农化巨头，与以崖州湾国家实验室、中国农业大学、先正达集团中国等为代表的中国团队，在模型架构、数据生态及战略演进逻辑上呈现出完全不同的路径特征。

国际农化巨头的竞争本质在于构建“种子+农化+数字化服务”的强耦合闭环，将AI深度嵌入从基因设计到田间决策的全生命周期。拜耳推出的Climate FieldView平台依托其庞大的数据底座，实现了品种表现的数字孪生模拟，利用大规模田间实时数据传至育种端大模型以动态优化育种方案[104]。科迪华则侧重于利用AI驱动的Precision Breeding平台，通过高维非线性模型解析复杂的“基因型-环境”交互作用[105]。

为验证此类闭环体系下AI赋能育种的实证效果，国际水稻研究所利用过去50年累积的连续栽培实验数据，构建了DL时序预测模型。该模型整合了近半个世纪的每日气象观测、土壤性状及产量记录，利用循环神经网络捕捉环境动态对产量的非线性影响[106]。结合iGEP框架，该模型成功识别出影响长期产量稳定的核心气候因子为夜间最低温度，并进一步证实夜温每升高1 ℃，特定感温品种灌浆速率将显著降低[106]。这一发现为培育耐高温的“绿色超级稻”提供了数字化判别标准，标志着水稻育种由单一基因选择转变为跨时空维度的环境适应性设计[107]。

针对全球小麦锈病的威胁，国际玉米小麦改良中心基于数百万个小麦基因型和历史抗病表现建立了大规模关联矩阵，利用多模态大模型和知识图谱的多跳推理，实现了从依赖田间表型向虚拟预测的战略转变[108]。育种者可利用LLMs的逻辑推理能力，从零散育种笔记及历史文献中提取关键抗性信息（如WGIE系统），并借助苗期DNA数据直接预测成株期的广谱抗性[42, 108]。这种虚拟筛选模式将育种周期缩短了数个生长季，使小麦育种进入了知识导航下的精准决策时代，有效克服了传统感病表型鉴定的滞后性。

与之形成鲜明对比的是，中国在智慧育种领域走出了“政策引导+有组织科研+开放共享”的特色路线。在种业振兴行动背景下，AI被定位为引领农业科研范式变革的引擎[109]。针对农业数据集少而分散的现状，中国采取了构建国家级种质资源大数据平台与开放育种生态的策略。例如，先正达集团中国利用其现代农业平台（Modern Agriculture Platform，MAP）覆盖的大规模耕地数据，通过与全球领先算法公司合作解码植物DNA语言，构建起集成基因型、表型、环境的全栈式预测体系[110]。

针对玉米育种中的表型获取瓶颈，研究者利用工业级高通量表型平台对4 000余株玉米种质进行了全生育期非破坏性动态监控，集成光谱相机和3D激光扫描仪获取厘米级的植株三维结构数据[111]。利用CNN对点云数据进行解析，实现了对叶片卷曲度、冠层温度变化的实时监测[112]。这种虚拟和现实相结合的监控体系，使AI能够识别受旱灾诱导的细微表型变异，并结合基因组信息成功定位多个抗旱关键QTL位点，显著提高了玉米在极端气候下的稳产性。

在大豆复杂性状遗传改良中，DL模型FE-WDNA（Feature Extraction for Whole DNA）的开发应用，标志着特征提取逻辑由原来的一维变异变成了现在的三维拓扑[113]。最近开发的SoyDNGP框架利用3D-CNN直接对原始DNA序列进行空间特征映射，解决了传统方法因SNP筛选导致的特征丢失问题[60]。通过残差网络捕捉序列中的长程依赖及上位性效应，SoyDNGP对大豆油脂含量、蛋白质产量等微效多基因调控性状展现出极强的泛化能力[60]。这种全栈式解析途径的预测精度显著优于传统的线性统计模型，为解析大豆遗传多样性提供了高效工具。

综上，国外大公司已经在数据厚度与全产业链闭环上形成了强大的护城河，而中国则依靠宏观政策、算力布局及开源生态的快速迭代寻找突破。为了更直观地呈现全球和国内智能设计育种基础模型的研发现状与竞争格局，本文系统梳理了当前在主粮作物性状预测、基因定向挖掘、表型特征解析及智慧育种决策中表现突出的代表性基础模型与平台。主要作物育种大模型的基本信息、模型类型、核心功能与项目来源如表1所示。这些垂直领域模型不仅展示了从单一文本检索向多模态跨媒介推理的技术演进，也勾勒出未来智慧育种系统朝着集约化、普惠化与端侧轻量化方向发展的产业全景底座。未来5—10年竞争焦点将不再局限于单一算法的优劣，而在于谁能更高效地将PB（Petabytes）级生物序列数据、万亿级气象/土壤时序数据与复杂的育种决策逻辑进行深度耦合。主粮育种的成功实践证明，育种家已具备在虚拟空间内预演并优化遗传增益的能力，通过构建具备生物学因果推理能力且可信任的智慧育种系统，将成为赢得全球种业科技博弈的关键。

表1 主要作物育种大模型

Table 1 Large-scale models for major crop breeding

名称	类型	主要功能描述	发布机构及时间
司农(Sinong)	通用农业领域垂直大模型	全学科知识检索与农业生产问题诊断	南京农业大学,2026
农业科学发现大模型	农业科研领域大模型	基因挖掘、性状预测、方案设计、种植系统智慧管理	中国农业科学院生物技术研究所/华为云,2025
神农大模型 3.0	农业垂直领域大模型	智慧育种、智慧种植、智慧养殖、遥感气象、智慧教育	中国农业大学,2025
AI Breeder for Crops (ABC)	集成化智慧育种平台	棉花一站式育种平台,并逐步拓展至水稻等多作物平台	浙江大学/北京大学,2025
西南种芯 V1	玉米智慧育种大模型	西南特色玉米的遗传演化分析、特异性状挖掘及改良	四川农业大学/博瑞迪,2025
天工开悟 2.0	农业大语言模型	病虫害预警、产量预测及生产要素的动态优化配置	哈尔滨工业大学,2025
作物育种大模型 (BreedingGPT)	育种专业辅助模型	科研问答,基因解析,文献检索,科研助手	北京大学现代农业研究院,2025
PlantGPT	大语言模型人工智能助手	植物功能基因组学智能问答、普及农业知识	华南农业大学,2025
SeedBench	多模态大模型测评基准	评估大模型在基因信息获取与分析、基因功能与调控机制分析、品种选育与农艺特性优化中的效用	上海人工智能实验室、崖州湾国家实验室,2025
CropGPT	生物大数据模型	协同应用AI、基因组编辑、多组学等前沿技术,丰富基础种质资源,提高育种精准	华中农业大学,2024
丰登(SeedLLM)	种业大语言模型	品种选育、栽培管理、种子经营、种业管理	上海人工智能实验室/崖州湾国家实验室/中国农业大学,2024
WBLM(Wheat Breeding LLM)	多模态育种大模型	融合基因型、表型、遥感、环境数据;性状预测与育种决策	康奈尔大学/浙江大学,2025
CropOS	工业AI育种平台	基因型-表型预测、候选品系筛选、目标性状设计	美国Benson Hill公司,2025
SEEDesign™	AI+基因编辑设计平台	in-silico性状设计、预测编辑位点、加速品种创制	美国农业科技公司(Inari),2025
PlantCaduceus	植物特异DNA基础模型	专门针对植物基因组训练;增强跨物种预测能力	开源植物基因组AI项目,2024
PLLaMA	植物科学基础开源大模型	植物学知识问答、实验设计辅助及育种决策咨询	加州大学圣巴巴拉分校/林肯大学等,2024
AgroNT	植物基因组基础模型 (DNA LLM)	多作物基因组预训练;调控元件识别、表达预测、功能变异优先级评估、跨物种迁移学习	InstaDeep/先正达,2024
Mineral	农业机器人+AI感知基础模型	高通量田间表型采集,作物个体级识别与监测,病害/营养状态检测,数据驱动育种与管理决策	Alphabet X(Google Moonshot)农业AI项目,2024
FarmVibes AI	多模态农业基础模型平台	融合卫星、气候、土壤、农机数据,作物生长预测与产量分析,精准播种与施肥决策,农田"what-if"模拟分析	微软,2022

5 挑战与展望

在AI辅助生物育种的浪潮下，作物育种虽已步入智能设计的全新纪元，但要实现全流程的工业化与自动化，仍面临深层技术瓶颈与系统性挑战。作物育种的本质是在复杂的基因型G×E交互作用中，寻求遗传增益的最大化。随着育种家对植物内部生理机能、外部气候波动及栽培措施耦合关系的理解不断深入，AI的核心价值正从单纯的预测工具转向系统优化。然而，这一转型涉及数据、算法、算力及伦理等多维度的系统工程。AI驱动育种面临的核心挑战与应对策略如表2所示，为我国在"十五五"种业振兴行动中重构智慧育种主权、筑牢主粮安全的数字边境提供了确定性的策略。

表2 AI驱动育种面临的挑战与应对策略

Table 2 Challenges and countermeasures in AI-driven breeding

维度	挑战	对策
数据	数据孤岛、标准化缺失	FAIR原则、联邦学习、国家级数据版权平台
算力	算力不对称、成本高	育种专用算力集群、LoRA量化/轻量化技术
算法	黑盒模型、缺少生物学因果	iGEP框架、知识引导的自监督学习、RAG技术
人才	复合型人才匮乏	农学+AI数据科学交叉学科培养

注：FAIR即可发现（Findable）、可访问（Accessible）、可互操作（Interoperable）、可重用（Reusable）。

5.1 智慧育种面临的多维挑战

1）数据孤岛、标准化困境与版权僵局。当前农业领域的数据集呈现少而分散的碎片化特征。种业巨头出于商业竞争考虑，使其核心种质库处于高度封闭状态，而公共科研机构的数据则因缺乏有效的共享机制而分布散乱。此外，非标准化问题严重阻碍了大规模跨环境数据集的构建，不同来源的数据在测量协议、元数据定义上存在显著差异，导致模型难以在统一的本体框架下进行有效训练[114, 115]。更深层次的掣肘在于高质量AI数据体系所遭遇的版权困境。随着训练数据量呈指数级增长，数据归属权、授权费用及潜在的侵权风险已成为垂直大模型发展的主要障碍。如何在保证数据要素合规流转的基础上，构建具备来源可追溯性的高质量农业语料库，已成为智慧育种迈向工业化阶段的首要瓶颈。

2）算力资源的极端不对称与高昂成本。算力是制约智慧育种普惠化的物质屏障。育种大模型的预训练和微调极度依赖高性能GPU服务器集群，对于中小型种企和公共科研机构而言，构建自主算力底座面临着巨大的财务压力。这种算力鸿沟可能引发种业科技的二次分化。此外，如何优化显存利用率和计算效率，实现大模型在无人农机端或者手持探测终端上的轻量化部署，仍是打通AI从实验室到田间应用这“最后一公里”的技术难题[93]。

3）算法的生物学因果逻辑缺失。农业应用场景极具复杂性，基于特定生态区训练的模型在面对气候变化造成的新环境时，泛化能力往往显著下降。DL模型虽然具备强大的特征学习能力，但因缺乏明确的生物学机理解释，导致育种家对AI预测结果存在信任危机[116, 117]。当前AI治理逻辑正经历从可解释向可信任的转变，强调在提升系统稳健性与安全性的同时，建立多维度的信任体系。因此，亟须将生物学先验知识融入算法中，利用iGEP等框架实现知识引导的数据驱动范式重构，将模型从单纯的概率联想转变为具备生物学合理性的因果推断[72]。

4）复合型人才的结构性匮乏与转化率难题。智慧育种的内核是生物技术与信息技术的深度融合。目前，尖端算法的开发与育种实践之间存在断层，缺乏既通晓遗传育种理论又精通DL算法的复合型人才。这导致大量模拟结果停留在虚拟实验阶段，难以转化为真实的田间遗传增益，形成了算法前沿性与田间生产实效性之间的脱节。

5.2 未来发展重点方向

1）构建标准化、开放共享的数据生态。制定统一的表型数据采集标准并践行FAIR原则，即可发现（Findable）、可访问（Accessible）、可互操作（Interoperable）、可重用（Reusable），是行业进步的基础[118]。针对数据隐私，应推广联邦学习等技术，在保护商业秘密的前提下实现联合建模[119]。未来应建立标准化的智慧育种准确性评价体系，通过构建公共基准数据集（Benchmark Datasets）和多环境验证群体，量化评价AI模型在不同G×E组合下的预测精度。同时，应强化数据主权意识，依托国家级人工智能数据版权一体化综合服务平台，确立数据确权机制，筑牢智慧育种的数据底座，守护国家粮食安全的数字边境。

2）集约化算力配置与轻量化路径探索。策略上应依托国家级算力枢纽建立育种专用算力集群，通过集约化配置解决中小型机构的算力缺口。技术上需进一步挖掘轻量化路径，利用参数高效微调、模型量化及知识蒸馏等技术手段，实现从云端巨模型向端侧轻模型的转化，赋能田间实时决策[93]。

3）深化AI与合成生物学的逻辑融合。育种的终极范式正从筛选变异转向创造变异。利用AI解析环境因子与栽培措施的交互逻辑，不仅能预测产量，更能反向优化栽培建议与品种选配策略。在此进程中，农业育种大模型需积极吸收基础生命科学领域的跨界突破。例如，Evo模型实现了从分子到

染色体尺度的长序列建模与设计，为重构作物复杂全基因组调控序列提供了宏观范式[120]；而DeepMind团队开发的AlphaGenome模型则显著提升了非编码区调控变异效应的预测精度，为解析作物微效多基因精密调控网络开辟了新路径[121]。将这些前沿模型的架构与作物特异性数据深度聚合，可更高效地辅助育种家设计最优代谢路径或人工蛋白结构，并辅以CRISPR-Cas等精准编辑工具，打破自然选择的遗传增益限制[122]。在此过程中，具备自主规划能力的育种智能体将成为主要驱动力，通过执行DBTL的虚拟循环[123]，实现育种范式由经验性筛选向工程化制造的本质飞跃。

4）完善法律伦理规范与人才培养体系。需针对AI设计的智能选育制定配套的审定法规和生物安全预警制度，确保技术发展与监管政策相适应[55, 124]。同时，应加速农学教育的数智化转型，将数据科学深入融入传统育种课程，培养新一代能够驾驭AI工具的数智科学家，为智慧育种的可持续发展提供智力支撑。

6 结 语

AI对育种的赋能绝非单一算法的叠加，而是集数据要素、底层架构、复合人才与制度逻辑的系统性范式变革。基于对当前技术演进的深度剖析，本文认为现代育种正经历从相关性预测向因果性设计的本质跨越：

1）从黑盒概率向生物学语义对齐的逻辑转向。未来育种大模型的生命力不在于参数规模的盲目扩张，而在于生物学先验知识与计算逻辑的深度对齐。面对多变的全球气候条件与复杂的栽培措施，AI的介入能够帮助育种家将分散的环境因子与植物生理响应进行语义耦合，使模型不仅能“算”出表型，更能解释基因在特定逆境下的表达逻辑。这种从概率联想向因果推断的转向，将是破解AI育种信任危机、实现精准评价的唯一途径。

2）构建虚拟设计-物理创制的智慧育种双环驱动模式。智慧育种的成熟形态应表现为一种高度集约的DT闭环。一方面是基于大模型和智能体的虚拟设计，在计算机内部完成海量遗传组合的in silico筛选与路径优化，解决传统育种中筛选评价体系效率低、周期长的顽疾；另一方面是基于CRISPR-Cas与高通量表型的物理创制，实现设计方案的精准落地与表型实时回传。这种双环驱动模式将极大压缩遗传增益的获取周期，使育种从发现演进为设计。

3）智慧育种主权的重构：数据确权与算力普惠。在全球种业竞争的宏观背景下，提出数据要素的流动性与算力的普惠性将决定国家种业安全的底线。在“十五五”规划与“种业振兴行动”的引领下，中国应通过构建国家级人工智能数据版权平台与育种专用算力集群，破解数据孤岛效应与技术垄断。这不仅是技术层面的优化，更是从制度层面筑牢国家粮食安全的数字边境，确保AI驱动的育种创新能够真正聚合各方资源，转化为实实在在的田间增益。

综上所述，智慧育种新时代的开启，标志着人类对作物进化控制权的深度收敛。这种BT与IT的跨界融合，不仅为应对极端气候与粮食安全挑战提供了终极工具，更将重塑全球农业科研的竞争格局。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 王术坤, 韩磊. 中国种业发展形势与国际比较[J]. 农业现代化研究, 2022, 43(5): 814-822.
WANG S K, HAN L. Development situation and international comparison of China's seed industry[J]. Research of Agricultural Modernization, 2022, 43(5): 814-822.
[2] TIAN Z X, WANG J W, LI J Y, et al. Designing future crops: Challenges and strategies for sustainable agriculture[J]. The Plant Journal, 2021, 105(5): 1165-1178.
[3] 种聪, 郭雨溪, 岳希明. 中国种业振兴: 发展历程、关键问题与机制构建[J]. 农业现代化研究, 2023, 44(2): 205-213.
CHONG C, GUO Y X, YUE X M. The revitalization of China's seed industry: Development process, main problems, and mechanism construction[J]. Research of Agricultural Modernization, 2023, 44(2): 205-213.
[4] GONAL B, SOMVEER, DOGGALLI G, et al. Exploring the future of plant breeding: Advancements and challenges[J]. International Journal of Plant & Soil Science, 2024, 35(24): 49-55.
[5] HALLAUER A R, DARRAH L L. Compendium of recurrent selection methods and their application[J]. Critical Reviews in Plant Sciences, 1985, 3(1): 1-33.
[6] BANSAL A, SRIVASTAVA P A. Transcriptomics to metabolomics: A network perspective for big data[M]// Applying Big Data Analytics in Bioinformatics and Medicine: IGI Global, 2018: 188-206.
[7] HICKEY L T, HAFEEZ A N, ROBINSON H, et al. Breeding crops to feed 10 billion[J]. Nature Biotechnology, 2019, 37(7): 744-754.
[8] AHMAR S, GILL RALI, JUNG K H, et al. Conventional and molecular techniques from simple breeding to speed breeding in crop plants: Recent advances and future outlook[J]. International Journal of Molecular Sciences, 2020, 21(7): 2590.
[9] LI H H, RASHEED A, HICKEY L T, et al. Fast-forwarding genetic gain[J]. Trends in Plant Science, 2018, 23(3): 184-186.
[10] ZHANG Y, HUANG G M, ZHAO Y X, et al. Revolutionizing crop breeding: Next-generation artificial intelligence and big data-driven intelligent design[J]. Engineering, 2025, 44: 245-255.
[11] 谷晓峰, 张立超, 李慧慧, 等. 农业生物智能设计育种[J]. 中国

农业科技导报, 2025, 27(12): 1-13.
GU X F, ZHANG L C, LI H H, et al. Agro-biological intelligent design breeding[J]. Journal of Agricultural Science and Technology, 2025, 27(12): 1-13.
[12] LI J J, XU M L, XIANG L R, et al. Foundation models in smart agriculture: Basics, opportunities, and challenges[J]. Computers and Electronics in Agriculture, 2024, 222: 109032.
[13] EFTEKHARI M, MA C, ORLOV Y L. Editorial: Applications of artificial intelligence, machine learning, and deep learning in plant breeding[J]. Frontiers in Plant Science, 2024, 15: 1420938.
[14] TZACHOR A, DEVARE M, RICHARDS C, et al. Large language models and agricultural extension services[J]. Nature Food, 2023, 4(11): 941-948.
[15] NAJAFABADI M Y. From text to traits: Exploring the role of large language models in plant breeding[J]. Frontiers in Plant Science, 2025, 16: 1583344.
[16] YOOSEFZADEH-NAJAFABADI M. Merging traditional practices and modern technology through computational plant breeding [J]. Plant Physiology, 2025, 199: kiaf355.
[17] DIAMOND J. Evolution, consequences and future of plant and animal domestication[J]. Nature, 2002, 418(6898): 700-707.
[18] LEE J, CHIN J H, AHN S N, et al. Brief history and perspectives on plant breeding[M]// Current Technologies in Plant Molecular Breeding. Dordrecht: Springer Netherlands, 2015: 1-14.
[19] MEYER R S, PURUGGANAN M D. Evolution of crop species: Genetics of domestication and diversification[J]. Nature Reviews Genetics, 2013, 14(12): 840-852.
[20] 简六梅, 肖英杰, 严建兵. 从头驯化: 作物品种设计与培育的新方向[J]. 遗传, 2023, 45(9): 741-753.
JIAN L M, XIAO Y J, YAN J B. *De novo* domestication: A new way for crop design and breeding[J]. Hereditas (Beijing), 2023, 45(9): 741-753.
[21] DOEBLEY J F, GAUT B S, SMITH B D. The molecular genetics of crop domestication[J]. Cell, 2006, 127(7): 1309-1321.
[22] YOOSEFZADEH NAJAFABADI M, HESAMI M, RAJCAN I. Unveiling the mysteries of non-Mendelian heredity in plant breeding[J]. Plants, 2023, 12(10): 1956.
[23] PINGALI P L. Green revolution: Impacts, limits, and the path ahead[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(31): 12302-12308.
[24] 张丰勇, 应晓平, 张健, 等. 半矮秆基因 sd1 调控水稻重要农艺性状的研究进展[J]. 中国水稻科学, 2025, 39(1): 24-32.
ZHANG F Y, YING X P, ZHANG J, et al. Progress in regulation of important agronomic traits by semi-dwarf gene sd1 in rice[J]. Chinese Journal of Rice Science, 2025, 39(1): 24-32.
[25] HINA A, ABBASI A, ARSHAD M, et al. Utilization of multi-omics approaches for crop improvement[M]. Hoboken: John Wiley & Sons, Ltd, 2024: 91-121.
[26] OLADOSU Y, RAFII M Y, ABDULLAH N, et al. Principle and application of plant mutagenesis in crop improvement: A review[J]. Biotechnology & Biotechnological Equipment, 2016, 30(1): 1-16.
[27] MEUWISSEN T E, HAYES B J, GODDARD M E. Prediction of total genetic value using genome-wide dense marker maps[J]. Genetics, 2001, 157(4): 1819-1829.
[28] CROSSA J, PÉREZ-RODRÍGUEZ P, CUEVAS J, et al. Genomic selection in plant breeding: Methods, models, and perspectives [J]. Trends in Plant Science, 2017, 22(11): 961-975.
[29] 方宣钧. 育种 5.0: 人工智能驱动的设计育种革命[J]. 分子植物育种(网络版), 2024, 22(3): 29-38.
FANG X J. Breeding 5.0: AI-driven revolution in designed plant breeding[J]. Molecular Plant Breeding (online), 2024, 22(3): 29-38.
[30] 徐云碧. 大数据与智慧育种[C]// 第二十届中国作物学会学术年会论文集. 长沙, 中国, 2023: 96.
[31] CHANDRA K, CHAND S, SAINI R P, et al. Smart breeding: Molecular interventions and advancements for crop improvement [M]. New York: Apple Academic Press, 2024.
[32] XU Y B, ZHANG X P, LI H H, et al. Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction[J]. Molecular Plant, 2022, 15(11): 1664-1695.
[33] CHEN K L, WANG Y P, ZHANG R, et al. CRISPR/cas genome editing and precision plant breeding in agriculture[J]. Annual Review of Plant Biology, 2019, 70: 667-697.
[34] ZHU H C, LI C, GAO C X. Applications of CRISPR–Cas in agriculture and plant biotechnology[J]. Nature Reviews Molecular Cell Biology, 2020, 21(11): 661-677.
[35] LI H H, LI X, ZHANG P, et al. Smart breeding platform: A web-based tool for high-throughput population genetics, phenomics, and genomic selection[J]. Molecular Plant, 2024, 17(5): 677-681.
[36] SUN L X, LAI M Y, GHOURI F, et al. Modern plant breeding techniques in crop improvement and genetic diversity: From molecular markers and gene editing to artificial intelligence—a critical review[J]. Plants, 2024, 13(19): 2676.
[37] YOOSEFZADEH NAJAFABADI M, LUKENS L, COSTA-NETO G. Editorial: Integrated omics approaches to accelerate plant improvement[J]. Frontiers in Plant Science, 2024, 15: 1397582.
[38] 周济, 陈佳玮, 沈利言, 等. 人工智能——推动植物研究发展的新动力[J]. 南京农业大学学报, 2022, 45(5): 1060-1071.
ZHOU J, CHEN J W, SHEN L Y, et al. Artificial intelligence: Advancing plant research beyond the state of the art[J]. Journal of Nanjing Agricultural University, 2022, 45(5): 1060-1071.
[39] LIU C, XU H, LI Z, et al. From map to blueprint: The plant pan-genome unraveling genetic mysteries and powering precision breeding[J]. Frontiers in Plant Science, 2025, 16: 1673637.
[40] LAM H Y I, ONG X E, MUTWIL M. Large language models in plant biology[J]. Trends in Plant Science, 2024, 29(10): 1145-1155.
[41] LIM S C, ITHARAJULA M, MØLLER M H, et al. PlantConnectome: A knowledge graph database encompassing >71, 000 plant articles[J]. The Plant Cell, 2025, 37(7): koaf169.
[42] WEI Y J, FAN J C. WGIE: Extraction of wheat germplasm resource information based on large language model[J]. Preprints, 2024: 2024111571.
[43] COSTA-NETO G, CRESPO-HERRERA L, FRADGLEY N, et al. Envirome-wide associations enhance multi-year genome-based prediction of historical wheat breeding data[J]. G3, 2023, 13(2): jkac313.
[44] STREICH J, ROMERO J, GAZOLLA J G F M, et al. Can exascale computing and explainable artificial intelligence applied to plant biology deliver on the United Nations sustainable development goals [J]. Current Opinion in Biotechnology, 2020, 61: 217-225.
[45] 杨莹莹, 郭嘉, 芦天罡, 等. 深度学习助力智慧农业高质量发展[J]. 农业工程, 2025, 15(12): 16-21.
YANG Y Y, GUO J, LU T G, et al. Deep learning empowers high-quality development of smart agriculture[J]. Agricultural Engineering, 2025, 15(12): 16-21.
[46] GAO T, SHENG W Y, ZHANG Z, et al. Greenhouse phenotyping measurement techniques and systems: A review[M]// Advanced Sensing and Robotics Technologies in Smart Agriculture. Singapore: Springer Nature Singapore, 2024: 43-59.
[47] XIANG L. High-throughput robotic plant phenotyping using 3D machine vision and deep neural networks[D]. Ames: Iowa State University, 2022.
[48] MAIMAITIJIANG M, SAGAN V, SIDIKE P, et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning[J]. Remote Sensing of Environment, 2020, 237: 111599.
[49] PURCELL W, NEUBAUER T, MALLINGER K. Digital Twins in agriculture: challenges and opportunities for environmental sustainability[J]. Current Opinion in Environmental Sustainability, 2023, 61: 101252.
[50] 李媛媛. 卷积神经网络优化及其在图像识别中的应用[D]. 沈阳: 沈阳工业大学, 2016.
LI Y Y. The optimization and application of convolution neural network for image recognition[D]. Shenyang: Shenyang University of Technology, 2016.
[51] VAN KLOMPENBURG T, KASSAHUN A, CATAL C. Crop

yield prediction using machine learning: A systematic literature review[J]. Computers and Electronics in Agriculture, 2020, 177: 105709.

[52] 吉文翰, 郑恒彪, 王迪, 等. 基于无人机影像和卷积神经网络的水稻育种材料产量预测研究[J]. 南京农业大学学报, 2025, 48(5): 1001-1012.
JI W H, ZHENG H B, WANG D, et al. Yield prediction in rice breeding materials using UAV-based images and convolutional neural network[J]. Journal of Nanjing Agricultural University, 2025, 48(5): 1001-1012.

[53] GUO W, FUKATSU T, NINOMIYA S. Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images[J]. Plant Methods, 2015, 11(1): 7.

[54] ABBAS A, JAIN S, GOUR M, et al. Tomato plant disease detection using transfer learning with C-GAN synthetic images[J]. Computers and Electronics in Agriculture, 2021, 187: 106279.

[55] MANSOOR S, KARUNATHILAKE E M B M, TUAN T T, et al. Genomics, phenomics, and machine learning in transforming plant research: Advancements and challenges[J]. Horticultural Plant Journal, 2025, 11(2): 486-503.

[56] 乔磊, 陈雷, 袁媛. 基于双意图建模和知识图谱扩散的水稻品种选育推荐方法[J]. 智慧农业(中英文), 2025, 7(2): 73-80.
QIAO L, CHEN L, YUAN Y. Bi-Intentional modeling and knowledge graph diffusion for rice variety selection and breeding recommendation[J]. Smart Agriculture, 2025, 7(2): 73-80.

[57] 鲍艳春, 石彩霞, 张传强, 等. 深度学习在基因组学中的研究进展[J]. 遗传, 2024, 46(9): 701-715.
BAO Y C, SHI C X, ZHANG C Q, et al. Progress on deep learning in genomics[J]. Hereditas, 2024, 46(9): 701-715.

[58] TOGNINALLI M, WANG X, KUCERA T, et al. Multi-modal deep learning improves grain yield prediction in wheat breeding by fusing genomics and phenomics[J]. Bioinformatics, 2023, 39(6): btad336.

[59] CLARK S A, VAN DER WERF J. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values[M]// Genome-Wide Association Studies and Genomic Prediction. Totowa: Humana Press, 2013: 321-330.

[60] GAO P F, ZHAO H N, LUO Z, et al. SoyDNGP: A web-accessible deep learning framework for genomic prediction in soybean breeding[J]. Briefings in Bioinformatics, 2023, 24(6): bbad349.

[61] DENG P C, LIU K N, ZHOU M X, et al. DPCformer: An interpretable deep learning model for genomic prediction in crops[EB/OL]. arXiv: 2510.08662, 2025.

[62] BACCUS S, BENGIO Y, BIRCH-SYKES C, et al. HyenaDNA: Long-range genomic sequence modeling at single nucleotide resolution[C]// Advances in Neural Information Processing Systems 36. New Orleans, Louisiana, USA. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023: 43177-43201.

[63] AVSEC Ž, AGARWAL V, VISENTIN D, et al. Effective gene expression prediction from sequence by integrating long-range interactions[J]. Nature Methods, 2021, 18(10): 1196-1203.

[64] 陈亦男, 史展. 改进DeepLabv3+模型的图像语义分割算法[J]. 黑龙江科技大学学报, 2026, 36(1): 163-168.
CHEN Y N, SHI Z. Image semantic segmentation algorithm based on improved DeepLabv3+model[J]. Journal of Heilongjiang University of Science and Technology, 2026, 36(1): 163-168.

[65] WU C L, ZHANG Y Y, YING Z W, et al. A transformer-based genomic prediction method fused with knowledge-guided module [J]. Briefings in Bioinformatics, 2023, 25: bbad438.

[66] CHEN J X, TAN C, ZHU M, et al. CropGS-Hub: A comprehensive database of genotype and phenotype resources for genomic prediction in major crops[J]. Nucleic Acids Research, 2024, 52(D1): D1519-D1529.

[67] BOULAHIA S Y, AMAMRA A, MADI M R, et al. Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition[J]. Machine Vision and Applications, 2021, 32(6): 121.

[68] YANG M G, CHEN S K, HUANG Z P, et al. Deep learning-enabled discovery and characterization of HKT genes in *Spartina alterniflora*[J]. The Plant Journal, 2023, 116(3): 690-705.

[69] STAHLSCHMIDT S R, ULFENBORG B, SYNNERGREN J. Multimodal deep learning for biomedical data fusion: A review [J]. Briefings in Bioinformatics, 2022, 23(2): bbab569.

[70] BALLARD J, WANG Z X, LI W R, et al. Deep learning-based approaches for multi-omics data integration and analysis[J]. BioData Mining, 2024, 17(1): 38.

[71] GUARRASI V, AKSU F, CARUSO C M, et al. A systematic review of intermediate fusion in multimodal deep learning for biomedical applications[J]. Image and Vision Computing, 2025, 158: 105509.

[72] WANG K L, ABID MALI, RASHEED A, et al. DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants[J]. Molecular Plant, 2023, 16(1): 279-293.

[73] 黄勃, 吴申奥, 王文广, 等. 图模互补: 知识图谱与大模型融合综述[J]. 武汉大学学报(理学版), 2024, 70(4): 397-412.
HUANG B, WU S A, WANG W G, et al. KG-LLM-MCom: A survey on integration of knowledge graph and large language model[J]. Journal of Wuhan University (Natural Science Edition), 2024, 70(4): 397-412.

[74] LARMANDE P, TODOROV K. AgroLD: A knowledge graph for the plant sciences[C]// The Semantic Web-ISWC 2021. Cham, Germany: Springer, 2021: 496-510.

[75] GUPTA P, ELSER J, HOOKS E, et al. Plant Reactome Knowledgebase: Empowering plant pathway exploration and OMICS data analysis[J]. Nucleic Acids Research, 2024, 52(D1): D1538-D1547.

[76] MENDOZA-REVILLA J, TROP E, GONZALEZ L, et al. A foundational large language model for edible plant genomes[J]. Communications Biology, 2024, 7: 835.

[77] LEVY B, XU Z H, ZHAO L Y, et al. FloraBERT: Cross-species transfer learning withattention-based neural networks for geneexpression prediction [EB/OL]. Research Square, (2022-08-30) [2026-02-05]. https://doi.org/10.21203/rs.3.rs-1927200/v1.

[78] 张子悦, 周欣智, 吕斌. 人工智能驱动的基因编辑技术及其应用[J]. 中国生物化学与分子生物学报, 2025, 41(4): 522-532.
ZHANG Z Y, ZHOU X Z, LÜ B. Harnessing artificial intelligence for advances in gene editing[J]. Chinese Journal of Biochemistry and Molecular Biology, 2025, 41(4): 522-532.

[79] RUFFOLO J A, NAYFACH S, GALLAGHER J, et al. Design of highly functional genome editors by modelling CRISPR-Cas sequences[J]. Nature, 2025, 645(8080): 518-525.

[80] 郭旺, 杨雨森, 吴华瑞, 等. 农业大模型: 关键技术、应用分析与发展方向[J]. 智慧农业(中英文), 2024, 6(2): 1-13.
GUO W, YANG Y S, WU H R, et al. Big models in agriculture: Key technologies, application and future directions[J]. Smart Agriculture, 2024, 6(2): 1-13.

[81] 冯在文, 许多, 田芳, 等. 设施农业智能决策大模型关键技术与发展方向[J]. 农业工程学报, 2025, 41(13): 50-61.
FENG Z W, XU D, TIAN F, et al. Key technologies and development trends of intelligent decision-making large models for facility agriculture[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(13): 50-61.

[82] BOMMASANI R, HUDSON DA, ADELI E, et al. On the opportunities and risks of foundation models[EB/OL]. arXiv: 2108.07258, 2021.

[83] THEODORIS C V, XIAO L, CHOPRA A, et al. Transfer learning enables predictions in network biology[J]. Nature, 2023, 618(7965): 616-624.

[84] ANTONELLO R J, VAIDYA A R, HUTH A G. Scaling laws for language encoding models in fMRI[J]. Advances in Neural Information Processing Systems, 2023, 36: 21895-21907.

[85] HOFFMANN J, BORGEAUD S, MENSCH A, et al. Training compute-optimal large language models[EB/OL]. arXiv: 2203.15556, 2022.

[86] WANG W X, HOSSEINI P, FEIZI S. Chain-of-defensive-thought: Structured reasoning elicits robustness in large language models against reference corruption[EB/OL]. arXiv: 2504.20769, 2025.

[87] WEI J, WANG X Z, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[EB/OL]. arXiv: 2201.11903, 2022.
[88] RAJBHANDARI S, RASLEY J, RUWASE O, et al. ZeRO: Memory optimizations toward training trillion parameter models[C]// SC20: International Conference for High Performance Computing, Networking, Storage and Analysis. Piscataway, New Jersey, USA. IEEE, 2020: 1-16.
[89] REZAYI S, LIU Z L, WU Z H, et al. AgriBERT: Knowledge-infused agricultural language models for matching food and nutrition[C]// Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2022: 5150-5156.
[90] YANG X J, GAO J F, XUE W X, et al. PLLaMa: An open-source large language model for plant science[EB/OL]. arXiv: 2401.01600, 2024.
[91] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[EB/OL]. arXiv: 2005.11401, 2020.
[92] LINEAR-PROBE A. Learning transferable visual models from natural language supervision[J]. Computer Science, 2021: 236776777.
[93] DING N, QIN Y J, YANG G, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models[J]. Nature Machine Intelligence, 2023, 5(3): 220-235.
[94] HU E J, SHEN Y L, WALLIS P, et al. LoRA: Low-rank adaptation of large language models[EB/OL]. arXiv: 2106.09685, 2021.
[95] VONDRACKOVA M, KOPCZYNSKI D, HOFFMANN N, et al. LORA, lipid over-representation analysis based on structural information[J]. Analytical Chemistry, 2023, 95(34): 12600-12604.
[96] ZHU Z D, SU Q D, DING Y Y, et al. LoRAFusion: Efficient LoRA fine-tuning for LLMs[EB/OL]. arXiv: 2510.00206, 2025.
[97] SAMANTA S, PRATIHAR S, CHATTERJI S. Precise and Quantitative Chlorosis Severity Assessment Framework (PQCSAF) using evolutionary superpixels[J]. Scientific Reports, 2025, 15: 37268.
[98] YING J, CHEN Z H, WANG Z F, et al. SeedBench: A multi-task benchmark for evaluating large language models in seed science [EB/OL]. arXiv: 2505.13220, 2025.
[99] 欧阳峥峥, 马毓聪, 寇远涛, 等. 中国农业大模型数据治理、核心技术与应用挑战[J]. 农业工程学报, 2025, 41(24): 153-162.
OUYANG Z Z, MA Y C, KOU Y T, et al. Data governance, core technologies, and application challenges of China's agricultural large models[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(24): 153-162.
[100] MCKENZIE P F, BERARDI A E, HOPKINS R. Delayed flowering phenology of red-flowering plants in response to hummingbird migration[J]. Current Biology, 2025, 35(9): 2175-2182.e3.
[101] ZIEGLER D M, STIENNON N, WU J, et al. Fine-tuning language models from human preferences[EB/OL]. arXiv: 1909.08593, 2019.
[102] KHOLOVÁ J, URBAN M O, COCK J, et al. In pursuit of a better world: Crop improvement and the CGIAR[J]. Journal of Experimental Botany, 2021, 72(14): 5158-5179.
[103] LYNAM J, BYERLEE D, MOOCK J L. The organizational challenge of international agricultural research: The fifty-year odyssey of the CGIAR[J]. Food Policy, 2024, 124(C): 102617.
[104] ANTLE J M, JONES J W, ROSENZWEIG C E. Next generation agricultural system data, models and knowledge products: Introduction[J]. Agricultural Systems, 2017, 155: 186-190.
[105] OLIVEIRA F A A, JONES J W, PAVAN W, et al. Incorporating a dynamic gene-based process module into a crop simulation model [J]. In Silico Plants, 2021, 3: diab011.
[106] PENG S B, HUANG J L, SHEEHY J E, et al. Rice yields decline with higher night temperature from global warming[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(27): 9971-9975.
[107] YU S B, ALI J, ZHOU S C, et al. From Green Super Rice to green agriculture: Reaping the promise of functional genomics research[J]. Molecular Plant, 2022, 15(1): 9-26.
[108] EDRISS V, JUMBO MDB, MAKUMBI D. Genomic prediction accuracy in CIMMYT's East African maize breeding program[C/OL]// Plant & Animal Genome (PAG) XXII. (2024-01-13) [2026-02-07]. https://pag.confex.com/pag/xxii/webprogram/Paper10657.html.
[109] 王可依. “粮农大讲堂” 第五期: 就我国种业振兴行动进行集中学习[J]. 世界农业, 2025(5): 138.
[110] 何玮萱, 刘琳, 胡博隆. 都市田园——互联网现代农业平台需求分析与开发推广研究[J]. 科技视界, 2018, 8(6): 28-29, 33.
HE W X, LIU L, HU B L. Urban idyllic - research on the demand analysis and development of Internet modern agriculture platform [J]. Science&Technology Vision, 2018, 8(6): 28-29, 33.
[111] SHEPARD N R, DESALVIO A J, ARIK M, et al. Deep learning-based high-throughput phenotyping of maize (Zea *mays* L.) tasseling from uas imagery across environments[EB/OL]. BioRxiv, 2024, https://doi.org/10.1101/2024.06.24.600506.
[112] SUN S P, LI C Y, PATERSON A H, et al. In-field high throughput phenotyping and cotton plant growth analysis using LiDAR[J]. Frontiers in Plant Science, 2018, 9: 16.
[113] WANG T, CUI Y P, SUN T, et al. A feature engineering method for whole-genome DNA sequence with nucleotide resolution[J]. International Journal of Molecular Sciences, 2025, 26(5): 2281.
[114] SELBY P, ABBELOOS R, BACKLUND J E, et al. BrAP: An application programming interface for plant breeding applications [J]. Bioinformatics, 2019, 35(20): 4147-4155.
[115] 万志前, 刘玉珍. 智慧农业背景下农业经营者数据权益保护困境与制度因应[J/OL]. 智慧农业(中英文), [2026-06-20]. Doi: 10.12133/j.smartag.SA202603023.
WAN Z Q, LIU Y Z. The protection dilemma and institutional responses of agricultural operators' data rights and interests in the context of smart agriculture[J]. Smart Agriculture, [2026-06-20]. Doi: 10.12133/j.smartag.SA202603023.
[116] 郭小东. 从“可解释”到“可信任”: 人工智能治理的逻辑重构[J]. 北京工业大学学报(社会科学版), 2025(6): 117-135.
GUO X D. From "explainable" to "trustworthy": A logical reconstruction of artificial intelligence governance[J]. Journal of Beijing University of Technology (Social Sciences Edition), 2025(6): 117-135.
[117] WALSH J J, MANGINA E, NEGRÃO S. Advancements in imaging sensors and AI for plant stress detection: A systematic literature review[J]. Plant Phenomics, 2024, 6: 153.
[118] WILKINSON M D, DUMONTIER M, AALBERSBERG I J, et al. The FAIR Guiding Principles for scientific data management and stewardship[J]. Scientific Data, 2016, 3: 160018.
[119] CHENG Y, LIU Y, CHEN T J, et al. Federated learning for privacy-preserving AI[J]. Communications of the ACM, 2020, 63(12): 33-36.
[120] NGUYEN E, POLI M, DURRANT M G, et al. Sequence modeling and design from molecular to genome scale with Evo[J]. Science, 2024, 386(6723): eado9336.
[121] AVSEC Ž, LATYSHEVA N, CHENG J, et al. Advancing regulatory variant effect prediction with AlphaGenome[J]. Nature, 2026, 649(8099): 1206-1218.
[122] ZHANG D L, XU F, WANG F H, et al. Synthetic biology and artificial intelligence in crop improvement[J]. Plant Communications, 2025, 6(2): 101220.
[123] JEON S, SOHN Y J, LEE H, et al. Recent advances in the Design-Build-Test-Learn (DBTL) cycle for systems metabolic engineering of *Corynebacterium glutamicum*[J]. Journal of Microbiology, 2025, 63(3): e2501021.
[124] POPA D. The ethics of using artificial intelligence in education [J]. International Journal of Communication Research, 2025, 15 (1): 75.

Artificial Intelligence Empowering Modern Agricultural Biological Breeding

XU Qiyu[1], ZHENG Bo[1, 2*], ZHONG Shangwei[1, 3]

(1. *Peking University Institute of Advanced Agricultural Sciences*, *Weifang* 261325, *China*; 2. *Weifang science and technology innovation promotion center*, *Weifang* 261071, *China*; 3. *College of Life Sciences*, *Peking University*, *Beijing* 100871, *China*)

Abstract: [Significance] The escalating complexity of genotype-phenotype-environment interactions and the explosive growth of multi-omics big data have necessitated a paradigm shift in crop breeding from empirical selection to intelligent design (Breeding 5.0). The aim of this paper is to systematically explore the underlying logic of the AI-driven crop breeding paradigm shift, comprehensively outline its generational evolution, analyze its core technical implementations in phenomics, genomics, multi-omics integration, and molecular design, and dissect how large agricultural foundation models reconstruct the entire seed industry workflow. **[Progress]** The historical evolution of crop breeding was first traced from 1.0 empirical domestication to 5.0 smart Breeding characterized by the deep integration of biotechnology (BT) and information technology (IT). In germplasm resource evaluation, deep learning algorithms enabled high-dimensional pattern recognition and unsupervised feature compression to unlock rare alleles from unannotated sequences, large language models (LLMs) like PlantConnectome and wheat germplasm information extraction (WGIE) leverage retrieval-augmented generation (RAG) to automatically construct structural knowledge graphs from unstructured historical literature, achieving predictive and dynamic germplasm evaluations. In high-throughput phenotyping, industrial platforms capture 3D point cloud and multi-spectral data at 0.1 mm resolution, while convolutional neural networks couple with the integrated genomic-enviromic prediction (iGEP) framework to build full-lifecycle digital twin crop models in virtual space. Regarding genomic prediction, the limitations of linear paradigms were dissected and cutting-edge deep learning architectures were highlighted: SoyDNGP applied a 3D-CNN to map chromosomal topology for complex soybean traits; DPCformer employed self-attention mechanisms to dynamically calculate environmental weights under multi-adversarial constraints; and HyenaDNA utilized long-convolution filters to bypass the $O(N^2)$ computational complexity limitation of standard Transformers, reducing it to $O(N \cdot log\ N)$ for chromosome-scale modeling. For multi-omics integration, intermediate fusion strategies were elucidated for their superior capacity to capture cross-layer biological compensatory pathways. In molecular design breeding, foundational plant language models were highlighted, such as AgroNT for zero-shot expression prediction and OpenCRISPR-1, the world's first de novo AI-generated genome editor built to capture underlying physico-chemical syntax. Furthermore, the emergence logic of major domestic and international agricultural foundation models was analyzed through the mathematical lenses of scaling laws, parameter-efficient fine-tuning, and multi-task evaluation benchmarks. Finally, empirical effectiveness was comprehensively evaluated through multinational success stories, including Bayer's Climate FieldView, IRRI's night-temperature thermal models for "Green Super Rice", and China's state-led breeding platforms for stress-resistant maize and high-yield soybean. **[Conclusions and Prospects]** Key structural challenges that smart breeding faces are thoroughly dissected: data silos and standardization dilemmas, extreme computational resource asymmetry and high training costs, the lack of biological causal logic in deep learning "black boxes", and the structural scarcity of interdisciplinary BT-IT talent. To overcome these bottlenecks, future research and policy efforts should focus on four pillars: 1) Establishing standardized open data ecosystems following FAIR (Findable, Accessible, Interoperable, Reusable) principles and promoting federated learning under a national AI data copyright integration platform to safeguard digital borders; 2) Exploring cloud-to-edge lightweight model deployment pathways through parameter-efficient fine-tuning, quantization, and knowledge distillation to empower real-time field-side decision-making and realize compute equity; 3) Deepening the mechanistic integration of AI and synthetic biology by absorbing breakthroughs from multi-scale modeling frameworks into the virtual "design-build-test-learn" cycle to break natural evolutionary thresholds; and 4) improving regulatory approval frameworks for AI-designed crops and modernizing agricultural education to cultivate a new generation of digital agronomists.

Keywords: smart breeding; artificial intelligence; breeding model; multi-omics fusion; high-throughput phenotyping; genomic selection; intelligent design breeding

Foundation items: National Natural Science Foundation of China (32300248); Natural Science Foundation of Shandong Province (ZR2022QC010)

First author: XU Qiyu, E-mail: xuqiyu@nwafu.edu.cn

***Corresponding author:** ZHENG Bo, E-mail: bo.zheng@pku-iaas.edu.cn