GENERAL STUDIES III

Genome Language Models and AI in Genomics – UPSC 2027

Faculty/Author: Faculty Editorial Board **Date:** 24 November 2026
Platform: Civil Service Gurukul Research Wing **Reading Time:** 5 Mins

Genome Language Models (gLMs) represent an emerging application of artificial intelligence that interprets DNA and RNA sequences by identifying complex biological patterns. Advanced models such as Evo 2, Nucleotide Transformer and DNABERT offer opportunities in genetic disease research, precision medicine, drug discovery, agriculture and biotechnology. India's GenomeIndia Project provides an important foundation for population-specific genomic research. This article examines the working principles, applications, scientific limitations, ethical concerns, biosecurity challenges and policy implications of genomic AI for UPSC Prelims and Mains 2027

CIVIL SERVICE GURUKUL
KNOWLEDGE | STRATEGY | SUCCESS
www.civilservicegurukul.com

UPSC CURRENT AFFAIRS | SCIENCE & TECHNOLOGY | BIOTECHNOLOGY

Artificial intelligence is moving beyond human language to explore the molecular information underlying living organisms. Genome Language Models are opening new avenues for understanding genetic variation, biological functions, disease mechanisms and the future of precision medicine.

24 November 2026 | Civil Service Gurukul | UPSC Prelims & Mains 2027

Why in the News?

Artificial intelligence-based Genome Language Models have emerged as an important research frontier at the intersection of biotechnology, computational biology and machine learning.

Research published in Nature in 2026 described Evo 2, a biological foundation model trained on approximately nine trillion DNA base pairs collected from diverse organisms.

Other developments, including GPN-Star and the Nucleotide Transformer family, demonstrate how artificial intelligence can help researchers investigate genetic variation, regulatory sequences and genome-wide biological functions.

These developments are significant because conventional genome sequencing produces enormous amounts of biological information, but interpreting the functional significance of that information remains a major scientific challenge.

For India, initiatives such as the GenomeIndia Project provide opportunities to combine genomic databases, artificial intelligence and biotechnology research for population-specific healthcare and scientific innovation.

EVO 2 TRAINING DATA

9 Trillion

DNA base pairs used for model training

EVO 2 CONTEXT WINDOW

1 Million

Tokens at single-nucleotide resolution

GENOMEINDIA MILESTONE

10,000+

Individuals represented through whole-genome sequencing

DNA BUILDING BLOCKS

A, T, G, C

Four principal nucleotide bases

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

In This Article

1. Why in the News?
2. UPSC Syllabus Relevance
3. Understanding DNA, Genes and Genomes
4. What Are Genome Language Models?
5. How Genome Language Models Work
6. GLMs vs Large Language Models
7. Major Genome Language Models
8. Applications of AI in Genomics
9. India and Genomic Research
10. Ethical and Biosecurity Concerns
11. Scientific and Technological Challenges
12. Way Forward
13. UPSC Prelims Revision Facts
14. UPSC Prelims Practice MCQs
15. UPSC Mains Practice Questions
16. UPSC Mains Model Answer
17. Frequently Asked Questions

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

UPSC CSE Syllabus Relevance

UPSC Paper	Relevant Syllabus Topics
Prelims GS Paper I	General Science, biotechnology, genetics and current affairs
Mains GS Paper II	Health services, government policies, public health and equitable access to technology
Mains GS Paper III	Artificial intelligence, biotechnology, scientific developments, intellectual property and technological applications
Mains GS Paper IV	Bioethics, privacy, informed consent, accountability and responsible innovation
Essay Paper	Science and society, technological progress, human dignity and ethical governance

Understanding DNA, Genes and Genomes

What Is DNA?

Deoxyribonucleic Acid, commonly known as DNA, is the principal hereditary material in cellular organisms. It stores biological information required for development, functioning and reproduction.

DNA consists of nucleotide units containing a sugar, a phosphate group and a nitrogenous base.

DNA Base	Full Name	Complementary Base
A	Adenine	Thymine
T	Thymine	Adenine
G	Guanine	Cytosine
C	Cytosine	Guanine

What Is a Gene?

A gene is a functional unit of hereditary information. Genes contain information used to produce functional RNA molecules or, in the case of protein-coding genes, proteins through RNA intermediates.

Genes influence various biological characteristics, although many human traits result from interactions involving multiple genes, regulatory mechanisms and environmental factors.

What Is a Genome?

A genome refers to the complete genetic material of an organism. It includes genes, regulatory sequences and other genetic regions.

A haploid human nuclear genome contains approximately three billion DNA base pairs. Humans also possess mitochondrial DNA.

What Is Genomics?

Genomics is the scientific study of genomes, including their structure, function, variation, regulation and evolution.

> **UPSC Prelims Conceptual Clarity:**
>
> Genetics generally focuses on heredity and individual genes, whereas genomics examines genetic material at a broader genome-wide level.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Central Dogma of Molecular Biology

The central dogma explains important pathways through which genetic information is transferred and expressed within biological systems.

DNA ? RNA ? Protein

Transcription converts genetic information in DNA into RNA. Translation uses messenger RNA information to produce proteins.

Reverse transcription, in which RNA serves as a template for DNA synthesis, is also an established biological process. Not all RNA molecules are translated into proteins.

Understanding these biological processes is essential for examining the role of artificial intelligence in genomics.

What Are Genome Language Models?

Genome Language Models, commonly abbreviated as gLMs or GLMs, are artificial intelligence models that learn patterns from genetic sequences.

They apply computational techniques inspired by natural language processing to biological information encoded in DNA and RNA.

Just as a conventional language model learns statistical relationships between words, a genomic language model learns relationships among nucleotides, sequence patterns and their surrounding genetic context.

The analogy between biological sequences and human language is useful for understanding the technology. However, DNA does not possess grammar or meaning in exactly the same sense as human language.

Genomic functions also depend on cellular environments, evolutionary history, molecular interactions and regulatory processes that may not be fully captured by DNA sequence alone.

Civil Service Gurukul Concept:

A Genome Language Model is an AI-based computational system trained to recognize patterns in genetic sequences and support predictions about biological functions, genetic variations and regulatory mechanisms.

How Do Genome Language Models Work?

Stage 1: Genomic Data Collection

Researchers assemble genetic sequence datasets from suitable genomic databases and scientific studies.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

The quality, diversity and accuracy of training datasets influence the reliability of the resulting model.

Stage 2: Tokenization

Genetic sequences are converted into computational units called tokens. Depending on the model, a token may represent an individual nucleotide or a short group of nucleotides.

Stage 3: Model Pretraining

The artificial intelligence model learns recurring statistical relationships from extensive genomic datasets, often using self-supervised learning.

Certain models learn by predicting missing sequence information, while others learn by predicting subsequent tokens.

Stage 4: Contextual Pattern Recognition

The model develops internal representations of sequence relationships and may recognize patterns associated with genes, regulatory regions and evolutionary conservation.

Stage 5: Application to Research Tasks

The pretrained model may be adapted to tasks such as genetic variant assessment, genome annotation and prediction of regulatory properties.

Stage 6: Scientific Validation

Researchers evaluate model predictions against independent datasets, established biological evidence and appropriate experimental observations.

AI predictions must not be treated as automatically established biological or clinical facts.

Important Artificial Intelligence Concepts Used in Genomics

AI Concept	Meaning in Genomic Research
Machine Learning	Learning statistical relationships from biological datasets
Deep Learning	Using multilayer neural networks to model complex sequence patterns
Transformer Architecture	Using attention mechanisms to represent relationships between sequence elements

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

AI Concept	Meaning in Genomic Research
Self-Supervised Learning	Learning from training data without requiring manually assigned labels for every example
Foundation Model	A broadly pretrained model that can be adapted to multiple research tasks
Tokenization	Representing genetic sequences as computational units
Fine-Tuning	Adapting a pretrained model to a more specific research task
Context Window	The amount of sequence information a model can consider within a given input context
Bioinformatics	The application of computational methods to biological information

Genome Language Models vs Large Language Models

Particulars	Genome Language Models	Text-Based Large Language Models
Primary Data	Genetic sequences	Human language and text
Basic Units	Nucleotides or nucleotide groups	Text tokens, words or subwords
Primary Objective	Learn biological sequence patterns	Learn linguistic and textual patterns
Typical Applications	Genome annotation and genetic variant research	Writing, translation, coding and question answering
Scientific Validation	Biological experiments, independent genomic data and clinical validation where relevant	Task-specific evaluation and factual verification
Major Ethical Concerns	Genetic privacy, discrimination and biological dual-use risks	Misinformation, privacy, bias and harmful misuse

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Major Genome Language Models

1. Evo 2

Evo 2 is a biological foundation model described in peer-reviewed research published in Nature in 2026.

Its training dataset comprises approximately nine trillion DNA base pairs spanning diverse forms of life. The model supports a context window of up to one million tokens at single-nucleotide resolution.

The research demonstrates advances in genomic pattern recognition, functional prediction and biological sequence modelling.

Evo 2 is a research technology. Its predictions do not independently establish clinical validity, biological safety or the feasibility of every predicted function.

2. Nucleotide Transformer

Nucleotide Transformer is a family of DNA foundation models developed to learn genomic sequence representations.

Research has examined models trained using diverse human and multispecies genome datasets for tasks associated with functional genomics and regulatory sequence analysis.

3. DNABERT

DNABERT applies language-model architecture to DNA sequence analysis.

It demonstrates how methods inspired by natural language processing can be adapted to identify biological patterns and support genome annotation tasks.

4. GPN and GPN-Star

These genomic modelling approaches incorporate evolutionary information to assess functional constraints and investigate the potential effects of genetic variations.

GPN-Star research published in 2026 highlights the importance of incorporating evolutionary relationships when evaluating genomic variation.

5. AlphaFold

AlphaFold is an important artificial intelligence system in structural biology. Its best-known application is predicting protein structures from sequence information.

Although closely related to AI-driven biological research, AlphaFold should not be classified simply as a DNA language model.

AI Model	Primary Research Focus
Evo 2	Large-scale genomic sequence modelling
DNABERT	DNA sequence representation and annotation
Nucleotide Transformer	Genomic foundation modelling and functional prediction
GPN-Star	Evolutionary constraints and genetic variant effects
AlphaFold	Protein structure prediction and related molecular modelling

Important UPSC Prelims Trap:

Genome sequencing, genome language modelling, protein structure prediction and genome editing are related but distinct scientific technologies.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Major Applications of Artificial Intelligence in Genomic Research

1. Genetic Disease Research

Genomic AI may help researchers assess whether particular genetic variations are likely to influence biological functions.

This could support research into inherited disorders, rare diseases and complex genetic conditions.

A computational prediction alone is not sufficient to establish a medical diagnosis.

2. Precision Medicine

Precision medicine seeks to account for differences in genetics, environment and lifestyle when developing disease prevention and treatment strategies.

AI-assisted genomic analysis may contribute to more individualized medical research and treatment selection after appropriate clinical validation.

3. Cancer Genomics

Cancer may involve genetic and regulatory alterations affecting normal cellular functions.

AI-based analysis can help researchers prioritize genetic changes for investigation and improve understanding of tumour biology.

4. Drug Discovery

Genomic information can help researchers identify biological pathways and investigate potential therapeutic targets.

Artificial intelligence may improve the efficiency of early-stage research, although laboratory studies, clinical trials and regulatory approvals remain essential.

5. Rare Disease Diagnosis Research

Genome sequencing can reveal numerous genetic variants whose clinical significance is uncertain.

AI models may help prioritize variants for further investigation by qualified genetics specialists.

6. Agricultural Biotechnology

AI-assisted genomic research may support crop breeding studies, agricultural biodiversity assessment and research into characteristics such as drought tolerance, nutritional quality and environmental resilience.

7. Biodiversity Conservation

Genomic analysis can help researchers understand evolutionary relationships, genetic diversity and population-level conservation priorities.

AI tools may assist in processing extensive biological datasets relevant to conservation research.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

8. Pharmacogenomics

Pharmacogenomics examines how genetic variation influences responses to medicines.

AI-assisted research may improve understanding of variation in drug response and support future treatment personalization where sufficient medical evidence exists.

9. Functional Genomics

Functional genomics investigates how genetic sequences contribute to biological functions and gene regulation.

Genome Language Models may support the identification of genomic regions requiring additional experimental investigation.

10. Evolutionary Biology

AI-assisted comparison of genetic information across organisms can support research into evolutionary conservation, biological diversity and the development of different characteristics over time.

Difference Between Genome Sequencing, Genome Editing and Genomic AI

Technology	Primary Function
Genome Sequencing	Determines the order of genetic sequence components
Genome Language Models	Computationally analyse and predict patterns in genetic information
Genome Editing	Introduces targeted changes to genetic material
CRISPR-Cas Technology	A family of technologies used for targeted genetic modification and other molecular applications
Gene Therapy	Uses genetic approaches to treat or prevent certain diseases
Protein Structure Prediction	Computationally estimates the three-dimensional structure of proteins

Conceptual Distinction:

A genome language model does not physically edit an organism's genetic material merely by analysing its sequence. Computational predictions and biological interventions are separate stages requiring different technologies and safeguards.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

India and the Future of Artificial Intelligence in Genomics

India possesses substantial opportunities for genomic research because of its population diversity, expanding biotechnology ecosystem, scientific institutions and growing digital infrastructure.

1. GenomeIndia Project

The GenomeIndia Project is a national genomic research initiative funded by the Department of Biotechnology under the Ministry of Science and Technology.

Launched in January 2020, the project aims to establish a comprehensive understanding of genetic variations across diverse Indian population groups.

The project achieved a major milestone through whole-genome sequencing of more than 10,000 individuals. Genomic data was dedicated to the research community in January 2025, with access governed through appropriate research arrangements.

Particular	GenomeIndia Project
Launch	January 2020
Funding Department	Department of Biotechnology
Ministry	Ministry of Science and Technology
Initial Sequencing Milestone	Over 10,000 individuals
Major Objective	Understanding genetic variation across Indian populations
Research Significance	Population genetics, disease research and precision medicine

2. Indian Biological Data Centre

The Indian Biological Data Centre, or IBDC, provides national infrastructure for the management and sharing of biological research data.

Such infrastructure can support genomic research, computational biology and responsible development of artificial intelligence-based research applications.

3. IndiGen Programme

The IndiGen initiative associated with the Council of Scientific and Industrial Research has contributed to understanding genetic variation among Indian individuals.

Population-specific genomic information can help address the limitations of research datasets that underrepresent Indian communities.

4. IndiaAI Mission

The IndiaAI Mission seeks to strengthen India's artificial intelligence ecosystem through computing infrastructure, research, innovation, datasets and skill development.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Its broader capabilities may support future AI applications in healthcare, biotechnology and scientific research.

5. BioE3 Policy

India's BioE3 Policy emphasizes Biotechnology for Economy, Environment and Employment.

It supports the development of high-performance biomanufacturing and biotechnology innovation.

AI-driven biological research may complement the wider objectives of this policy.

Indian Initiative	Major Relevance
GenomeIndia	Indian population genomic diversity
Indian Biological Data Centre	Biological data infrastructure
IndiGen	Population-level genome research
IndiaAI Mission	Artificial intelligence ecosystem and computing capacity
BioE3 Policy	Biotechnology innovation and biomanufacturing
Biotechnology Industry Research Assistance Council	Support for biotechnology research and entrepreneurship

These initiatives represent complementary areas of national scientific capacity. Their inclusion does not imply that all programmes are currently integrated into a single Genome Language Model project.

Why India Needs Population-Specific Genomic Research

Genetic variation differs across populations and individuals. Artificial intelligence models trained predominantly on datasets from other regions may not perform equally well for every Indian population.

Population-representative research can contribute to reducing this gap while recognizing that ancestry is only one factor affecting health outcomes.

- Better understanding of Indian genetic diversity.
- Research into inherited and rare diseases.
- Improved representation in international genomic studies.
- Population-specific evaluation of medical AI models.
- Development of domestic biotechnology research capabilities.
- Potential improvements in precision healthcare.
- Strengthening scientific data infrastructure.
- Promotion of interdisciplinary biotechnology education.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Ethical Concerns Associated with Genome Language Models

1. Genetic Data Privacy

Genomic information may reveal sensitive details concerning individuals, their biological relatives and inherited characteristics.

Removing names and conventional identifiers may not eliminate every possibility of re-identification.

2. Informed Consent

Individuals participating in genomic studies should receive understandable information about data collection, research purposes, storage, sharing and potential secondary uses.

3. Genetic Discrimination

Inappropriate use of genetic information may create risks of discrimination in employment, insurance, healthcare or other social contexts.

4. Algorithmic Bias

Underrepresentation of particular population groups may reduce the accuracy and fairness of AI-assisted genomic analysis.

5. Data Ownership and Benefit Sharing

Genomic research raises questions concerning control over biological information, commercial use, intellectual property and equitable sharing of scientific benefits.

6. Commercialization of Genetic Information

Commercial research involving genomic datasets should operate within appropriate consent, privacy and governance frameworks.

7. Biosecurity and Dual-Use Concerns

Advanced biological AI can support beneficial scientific research, but certain capabilities may also create risks of harmful misuse.

Responsible governance requires proportionate access controls, expert oversight, safety evaluation and appropriate safeguards against misuse.

8. Clinical Accountability

Medical decisions should not rely solely on unvalidated AI predictions. Healthcare professionals, research institutions and technology developers must establish appropriate accountability and review mechanisms.

UPSC Ethics Perspective:

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

> Genomic AI involves a balance between scientific freedom, individual privacy, public health, equitable access, human dignity and collective safety. Technological capability alone cannot determine whether a particular application is ethically justified.

Legal and Regulatory Dimensions in India

Digital Personal Data Protection Act, 2023

The Digital Personal Data Protection Act establishes a framework for processing digital personal data in India.

Identifiable genomic information may fall within the scope of personal data, depending on the circumstances and applicable legal provisions.

Research institutions should examine applicable commencement provisions, rules, exemptions and data protection obligations before processing such information.

Right to Privacy

In Justice K.S. Puttaswamy v. Union of India, the Supreme Court recognized privacy as a fundamental right protected under the Constitution.

This constitutional principle is particularly relevant to the responsible management of sensitive biological and personal information.

ICMR Ethical Guidelines

The Indian Council of Medical Research provides ethical guidance for biomedical and health research involving human participants.

Genomic studies should consider applicable requirements concerning informed consent, ethics committee oversight, confidentiality and participant protection.

Intellectual Property Rights

AI-driven biotechnology raises questions concerning patentability, ownership of inventions, access to research datasets and the commercial exploitation of scientific discoveries.

These matters must be assessed under applicable intellectual property laws rather than assuming that naturally occurring genetic information can automatically be patented.

Scientific and Technological Challenges

Challenge	Explanation

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Complex Genome Regulation	Biological functions depend on interactions extending beyond individual genetic sequences
Data Quality	Errors or incomplete datasets can reduce prediction accuracy
Population Bias	Unrepresentative datasets may limit performance across different populations
High Computational Requirements	Large AI models may require extensive computing and energy resources
Model Interpretability	Understanding why an AI model makes a prediction can be difficult
Long-Range Interactions	Relationships between distant genomic regions remain challenging to predict reliably
Clinical Validation	Research predictions must be independently evaluated before medical application
Environmental Effects	DNA sequences alone cannot explain all influences on health and biological characteristics
Epigenetic Factors	Gene regulation may depend on molecular changes and cellular conditions not captured by sequence alone
Biosecurity	Advanced biological modelling requires responsible safeguards

Can Artificial Intelligence Completely Decode the Human Genome?

Genome Language Models represent significant advances in computational biology, but they do not provide a complete understanding of all biological functions.

The relationship between genetic sequence and biological outcome is influenced by gene regulation, protein interactions, epigenetic mechanisms, environmental conditions and developmental processes.

A large context window or extensive training dataset does not necessarily guarantee accurate predictions across every scientific task.

Traditional statistical methods, evolutionary models and specialized biological tools may outperform genomic foundation models for particular applications.

Scientific Limitation:

Artificial intelligence identifies statistical relationships and generates predictions. Experimental research remains necessary to establish biological mechanisms, causation, clinical effectiveness and safety.

Genome Language Models and Sustainable Development Goals

Sustainable Development Goal	Potential Contribution

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

SDG 2 – Zero Hunger	Research supporting resilient crops and sustainable agriculture
SDG 3 – Good Health and Well-Being	Precision medicine, disease research and healthcare innovation
SDG 4 – Quality Education	Interdisciplinary training in biotechnology and artificial intelligence
SDG 9 – Industry, Innovation and Infrastructure	Scientific infrastructure and biotechnology innovation
SDG 10 – Reduced Inequalities	Improving representation and equitable access to genomic research
SDG 15 – Life on Land	Genetic diversity research and biodiversity conservation

Way Forward for India

1. Strengthen Representative Genomic Databases

India should expand ethically governed genomic research covering diverse population groups and address gaps in scientific representation.

2. Develop Indigenous AI Capabilities

Investment in domestic computing infrastructure, biotechnology research and artificial intelligence can strengthen national scientific capacity.

3. Promote Interdisciplinary Education

Universities and research institutions should encourage collaboration among geneticists, doctors, computer scientists, statisticians, ethicists and public policy experts.

4. Establish Responsible Data Governance

Genomic datasets should be managed using appropriate consent arrangements, privacy protections, cybersecurity measures and access controls.

5. Ensure Independent Scientific Validation

AI-based biological predictions should be evaluated using independent datasets and appropriate scientific or clinical research methods.

6. Strengthen Biosecurity Governance

Research institutions should incorporate safety assessments, responsible access policies and expert oversight into genomic AI development.

7. Promote Public-Private Research Partnerships

Collaboration between government institutions, universities, biotechnology companies and startups can support responsible scientific innovation.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

8. Ensure Equitable Access

The benefits of precision medicine and genomic research should not remain restricted to economically privileged groups or major urban centres.

9. Encourage International Cooperation

International scientific collaboration can support research reproducibility, responsible data sharing and common approaches to biological AI safety.

10. Strengthen Public Awareness

Citizens should receive accurate information about genomic research, its benefits, limitations and potential implications for privacy and healthcare.

UPSC Prelims 2027: Important Facts for Revision

1. DNA stands for Deoxyribonucleic Acid.
2. RNA stands for Ribonucleic Acid.
3. The four principal DNA bases are Adenine, Thymine, Guanine and Cytosine.
4. RNA typically contains Uracil instead of Thymine.
5. A haploid human nuclear genome contains approximately three billion base pairs.
6. A gene may encode a protein or a functional RNA molecule.
7. Genomics studies genetic information at a genome-wide level.
8. Genome Language Models learn statistical patterns from genetic sequences.
9. Self-supervised learning is commonly used in genomic foundation models.
10. Evo 2 was described in Nature in 2026.
11. Evo 2 was trained on approximately nine trillion DNA base pairs.
12. Evo 2 supports a context window of up to one million tokens.
13. DNABERT applies language modelling techniques to DNA sequences.
14. AlphaFold is primarily associated with protein structure prediction.
15. CRISPR-Cas technologies can be used for targeted genetic modification.
16. The GenomeIndia Project was launched in January 2020.
17. GenomeIndia is funded by the Department of Biotechnology.
18. GenomeIndia achieved whole-genome sequencing of more than 10,000 individuals.
19. The Indian Biological Data Centre supports biological data infrastructure.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

20. BioE3 stands for Biotechnology for Economy, Environment and Employment.
21. Pharmacogenomics studies how genetic variations influence drug responses.
22. Genomic AI predictions require independent scientific validation.

UPSC Prelims Practice Questions

Question 1

With reference to Genome Language Models, consider the following statements:

1. They can learn statistical patterns from DNA sequences.

2. They may support genetic variant effect prediction.

3. They directly modify the genetic material of living organisms during computational analysis.

Which of the statements given above are correct?

(a) 1 only
(b) 1 and 2 only
(c) 2 and 3 only
(d) 1, 2 and 3

View Answer and Explanation

Correct Answer: (b) 1 and 2 only

Statements 1 and 2 are correct because genomic AI models learn sequence relationships and may support functional predictions. Statement 3 is incorrect because computational modelling does not itself physically modify DNA.

Question 2

With reference to the artificial intelligence model Evo 2, consider the following statements:

1. It is a biological foundation model.

2. It was trained on approximately nine trillion DNA base pairs.

3. It was trained exclusively on the human genome.

Which of the statements given above are correct?

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

(a) 1 and 2 only
(b) 2 and 3 only
(c) 1 and 3 only
(d) 1, 2 and 3

View Answer and Explanation

Correct Answer: (a) 1 and 2 only

Evo 2 is a biological foundation model trained using approximately nine trillion DNA base pairs. Its dataset spans diverse organisms rather than being restricted to humans.

Question 3

The GenomeIndia Project is funded by which of the following departments?

(a) Department of Atomic Energy
(b) Department of Telecommunications
(c) Department of Biotechnology
(d) Department of Space

View Answer and Explanation

Correct Answer: (c) Department of Biotechnology

GenomeIndia is funded by the Department of Biotechnology under the Ministry of Science and Technology.

Question 4

Consider the following pairs:

Technology	Primary Application
1. AlphaFold	Protein structure prediction
2. DNABERT	DNA sequence modelling
3. CRISPR-Cas9	Targeted genome editing

How many of the above pairs are correctly matched?

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

(a) Only one
(b) Only two
(c) All three
(d) None

View Answer and Explanation

Correct Answer: (c) All three

AlphaFold is associated with protein structure prediction, DNABERT applies language modelling to DNA, and CRISPR-Cas9 can be used for targeted genome editing.

Question 5

Which one of the following nitrogenous bases is typically found in RNA instead of Thymine?

(a) Adenine
(b) Guanine
(c) Cytosine
(d) Uracil

View Answer and Explanation

Correct Answer: (d) Uracil

RNA typically contains Adenine, Uracil, Guanine and Cytosine, whereas DNA contains Thymine instead of Uracil.

Question 6

Consider the following statements regarding precision medicine:

1. It may consider genetic variation among individuals.

2. It may incorporate environmental and lifestyle information.

3. It eliminates the need for clinical validation of treatment approaches.

Which of the statements given above are correct?

(a) 1 and 2 only
(b) 2 and 3 only

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

(c) 1 and 3 only

(d) 1, 2 and 3

View Answer and Explanation

Correct Answer: (a) 1 and 2 only

Precision medicine considers individual variation, including genetic, environmental and lifestyle factors. It does not eliminate clinical evidence or validation requirements.

Question 7

The term pharmacogenomics refers to the study of:

(a) The geographical distribution of medicinal plants

(b) The influence of genetic variation on drug responses

(c) The industrial manufacture of pharmaceutical packaging

(d) The physical structure of chromosomes alone

View Answer and Explanation

Correct Answer: (b)

Pharmacogenomics examines how genetic differences influence an individual's response to medicines.

Question 8

Which of the following are ethical concerns associated with artificial intelligence in genomic research?

1. Genetic discrimination.

2. Unauthorized use of genetic information.

3. Population representation bias.

4. Biological dual-use risks.

Select the correct answer using the code given below:

(a) 1 and 2 only

(b) 2 and 3 only

(c) 1, 3 and 4 only

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

(d) 1, 2, 3 and 4

View Answer and Explanation

Correct Answer: (d) 1, 2, 3 and 4

All four represent important ethical or governance concerns in genomic AI research.

Question 9

The BioE3 Policy of the Government of India primarily refers to biotechnology for:

(a) Education, Engineering and Energy
(b) Economy, Environment and Employment
(c) Ecology, Electronics and Exports
(d) Evolution, Entrepreneurship and Equity

View Answer and Explanation

Correct Answer: (b) Economy, Environment and Employment

The BioE3 Policy focuses on biotechnology for economic development, environmental sustainability and employment generation.

Question 10

With reference to genomic artificial intelligence, consider the following statements:

1. DNA sequence information alone may not explain every biological characteristic.

2. Epigenetic and environmental factors can influence biological outcomes.

3. Increasing model size guarantees accurate predictions for all genomic research tasks.

Which of the statements given above are correct?

(a) 1 only
(b) 1 and 2 only
(c) 2 and 3 only
(d) 1, 2 and 3

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

View Answer and Explanation

Correct Answer: (b) 1 and 2 only

Biological outcomes are influenced by several interacting genetic and non-genetic factors. Larger AI models do not guarantee superior accuracy across every research task.

UPSC Mains Practice Questions

Question 1 – GS Paper III

Genome Language Models represent a convergence of artificial intelligence and biotechnology. Discuss their applications, scientific limitations and potential significance for India.

Marks: 15 | **Word Limit:** 250 words

Question 2 – GS Paper III

Examine the role of the Genome India Project in advancing population-specific genomic research and precision medicine.

Marks: 10 | **Word Limit:** 150 words

Question 3 – GS Paper IV

Discuss the ethical concerns associated with artificial intelligence-based genomic research. How can scientific innovation be balanced with genomic privacy and human dignity?

Marks: 15 | **Word Limit:** 250 words

Question 4 – GS Paper II

Evaluate the opportunities and challenges associated with integrating precision medicine and genomic artificial intelligence into India's public healthcare system.

Marks: 15 | **Word Limit:** 250 words

Question 5 – GS Paper III

Distinguish between genome sequencing, Genome Language Models and genome editing. Examine their complementary roles in modern biotechnology.

Marks: 10 | **Word Limit:** 150 words

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

UPSC Mains Model Answer

Question

Genome Language Models represent a convergence of artificial intelligence and biotechnology. Discuss their applications, scientific limitations and potential significance for India.

GS Paper III | 15 Marks | 250 Words

Introduction

Genome Language Models are artificial intelligence systems trained to recognise patterns in genetic sequences. Recent developments such as Evo 2 demonstrate their potential to advance computational biology and improve understanding of complex genomic information.

Major Applications

- **Healthcare:** Supporting research into inherited disorders, cancer and rare diseases.
- **Precision medicine:** Investigating genetic variation relevant to individualized treatment.
- **Drug discovery:** Assisting identification and prioritization of potential therapeutic targets.
- **Agriculture:** Supporting genomic research relevant to resilient and nutritious crops.
- **Conservation:** Improving understanding of genetic diversity and evolutionary relationships.

Significance for India

The GenomeIndia Project, Indian Biological Data Centre, IndiaAI Mission and BioE3 Policy provide complementary foundations for strengthening domestic genomic research, biotechnology innovation and computational capabilities.

Challenges

- Genetic privacy and informed consent.
- Population representation and algorithmic bias.
- High computational requirements.
- Limited model interpretability.
- Biological dual-use and biosecurity concerns.
- Need for independent experimental and clinical validation.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Way Forward

India should promote representative genomic databases, interdisciplinary education, indigenous AI development and responsible data governance. Independent scientific evaluation, ethical oversight and international cooperation are necessary to ensure safe and equitable innovation.

Conclusion

Genome Language Models offer significant scientific opportunities but cannot substitute for established biological and clinical evidence. A balanced governance framework integrating technological innovation, scientific accountability and human dignity can help India realize the benefits of genomic AI responsibly.

Important Keywords for UPSC Mains Answer Writing

Genomic Artificial Intelligence | Computational Biology | Precision Medicine | Population Genomics | Self-Supervised Learning | Biological Foundation Models | Genetic Diversity | Functional Genomics | Pharmacogenomics | Responsible Innovation | Genetic Data Sovereignty | Informed Consent | Algorithmic Fairness | Biosecurity | Dual-Use Technology | Scientific Validation | Digital Health | Inclusive Biotechnology | Data Governance | Ethical Artificial Intelligence

Frequently Asked Questions

1. What are Genome Language Models?

Genome Language Models are AI systems that learn patterns from DNA or RNA sequences to support biological research and functional predictions.

2. How are Genome Language Models different from ChatGPT?

Text-based language models primarily learn from human language, whereas Genome Language Models primarily learn relationships within biological sequences.

3. What is Evo 2?

Evo 2 is a biological foundation model described in scientific research published in 2026. It was trained on approximately nine trillion DNA base pairs.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

4. What is the context window of Evo 2?

Evo 2 supports a context window of up to one million tokens at single-nucleotide resolution.

5. What is the GenomeIndia Project?

GenomeIndia is a national initiative funded by the Department of Biotechnology to study genetic diversity across Indian population groups.

6. When was the GenomeIndia Project launched?

The project was formally launched in January 2020.

7. How many individuals were included in the initial GenomeIndia sequencing milestone?

The project achieved whole-genome sequencing of more than 10,000 individuals.

8. Can Genome Language Models diagnose diseases independently?

No. They may support research and clinical interpretation, but medical diagnoses require appropriate evidence and qualified professional assessment.

9. Are genome sequencing and genome editing the same?

No. Genome sequencing reads genetic information, whereas genome editing introduces changes to genetic material.

10. What is pharmacogenomics?

Pharmacogenomics studies the relationship between genetic variation and individual responses to medicines.

11. Is AlphaFold a DNA language model?

AlphaFold is primarily associated with protein structure prediction and related molecular modelling rather than being a DNA language model.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

12. What are the major ethical concerns associated with genomic AI?

Major concerns include genetic privacy, informed consent, discrimination, population bias, data ownership and biological misuse risks.

13. Can AI replace conventional biological research?

No. AI can complement scientific research, but independent experiments and appropriate validation remain necessary.

14. How can Genome Language Models benefit agriculture?

They may support research into crop genetics, biodiversity, environmental resilience and other agriculturally relevant characteristics.

15. Why are Genome Language Models important for UPSC 2027?

The topic connects biotechnology, artificial intelligence, healthcare, scientific innovation, data privacy, bioethics and responsible technology governance.

Conclusion

Genome Language Models represent an emerging frontier in the convergence of artificial intelligence and biological sciences.

By analysing patterns within extensive genetic datasets, these technologies may support advances in disease research, precision medicine, agricultural biotechnology, biodiversity conservation and scientific innovation.

India's growing genomic databases and biotechnology ecosystem create important opportunities for responsible participation in this scientific transformation.

However, the development of genomic artificial intelligence must be accompanied by strong data protection, equitable representation, scientific validation, ethical oversight and biosecurity governance.

The future of genomic research should be guided not merely by computational capabilities, but also by scientific integrity, public welfare and respect for human dignity.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

Prepare for UPSC CSE 2027 with Civil Service Gurukul

Explore Our Courses | Contact Our Academic Team | WhatsApp for Guidance | Email Us

Academic Disclaimer:

This article has been prepared for educational purposes and UPSC Civil Services Examination preparation. It reflects scientific and policy information reviewed as of 24 November 2026.

Potential applications of Genome Language Models should not be interpreted as established clinical treatments or guaranteed scientific outcomes. Independent validation and appropriate regulatory and ethical approvals remain necessary.

The practice questions and model answers are independently prepared educational material and are not represented as official UPSC previous-year questions.

CIVIL SERVICE GURUKUL
www.civilservicegurukul.com

CIVIL SERVICE GURUKUL

Empowering Aspirants | Building Future Civil Servants

CIVIL SERVICE GURUKUL

UPSC Preparation • Current Affairs • Mentorship • Test Series

Website: www.civilservicegurukul.com • Email: info@civilservicegurukul.com • Mobile: +91 9911020626