

Delfín Ortega-Sánchez y  
Alexander López-Padrón  
(eds.)

# Educación y sociedad: claves interdisciplinarias



# **Educación y sociedad: claves interdisciplinarias**

Delfín Ortega-Sánchez y Alexander López-Padrón (eds.)

**Octaedro**   
Editorial

COLECCIÓN: Universidad

TÍTULO: *Educación y sociedad: claves interdisciplinares*

EDICIÓN:

Delfín Ortega-Sánchez  
Alexander López-Padrón (eds.)

COMITÉ CIENTÍFICO INTERNACIONAL (edición de la obra):

Prof. Dr. Enric Bou, Università Ca' Foscari Venezia  
Prof. Dr. Julio Cabero Almenara, Universidad de Sevilla  
Prof. Dr. Miguel Cazorla Quevedo, Universidad de Alicante  
Prof. Dr. Antonio Cortijo, University of California at Santa Barbara  
Prof. Dra. José María Esteve Faubel, Universidad de Alicante  
Prof. Dr. Massimiliano Fiorucci, Università degli studi Roma Tre  
Prof. Dra. Carolina Flores Lueg, Universidad del Bío-Bío  
Prof. Dra. Mariana González Boluda, University of Leicester  
Prof. Dr. Alexander López Padrón, Universidad Técnica de Manabí  
Prof. Dr. Santiago Mengual Andrés, Universitat de València  
Prof. Dra. María Paz Prendes Espinosa, Universidad de Murcia  
Prof. Dra. Rozalya Sasor, Jagiellonian University in Kraków  
Prof. Dr. Fabrizio Manuel Sirignano, Università degli Studi Suor Orsola Benincasa  
Prof. Dra. Maria Stefanie Vasquez Peñafiel, Escuela Politécnica Nacional (Ecuador)

En este libro se recogen únicamente las aportaciones que han superado un riguroso proceso de selección y evaluación (*double blind peer review process*) según los siguientes criterios de evaluación: calidad del texto enviado, novedad y pertinencia del tema, originalidad de la propuesta, fundamentación bibliográfica y rigor científico.

Primera edición: noviembre de 2023

© De la edición: Delfin Ortega-Sánchez y Alexander López-Padrón

© Del texto: Las autoras y autores

© De esta edición:

Ediciones OCTAEDRO, S.L.  
C/ Bailén, 5 – 08010 Barcelona  
Tel.: 93 246 40 02 – Fax: 93 231 18 68  
www.octaedro.com – octaedro@octaedro.com

ISBN: 978-84-10054-35-6

Producción: Ediciones Octaedro

Esta publicación está sujeta a la Licencia Internacional Pública de Atribución/Reconocimiento-NoComercial 4.0 de Creative Commons. Puede consultar las condiciones de esta licencia si accede a: <https://creativecommons.org/licenses/by-nc/4.0/>.

NOTA EDITORIAL: Las opiniones y contenidos de los textos publicados en esta obra son de responsabilidad exclusiva de los autores.

Publicación en *Open Access* – Acceso abierto

# AI in Education: An Innovative Approach to Teaching Applied Bioinformatics with BioPython for Genomic Medicine

---

Juan Carlos Olabe Basogain

*Christian Brothers University*

Xabier Basogain Olabe  
Miguel Ángel Olabe Basogain

*University of the Basque Country*

**Abstract:** This chapter examines an innovative approach to teaching applied bioinformatics for genomic medicine using BioPython, targeted at high school or early college students. Recognizing the growing need for bioinformatics skills in the era of genomic medicine, we designed this course to provide a comprehensive introduction to programming for biological data analysis. The distinctive aspect of this educational initiative is the integration of the GPT-4, an advanced AI language model, into the course development and teaching process. The model assists in lesson planning, generating context-relevant examples, providing feedback on students' code, and facilitating collaborative problem-solving. It also aids in debugging, algorithm understanding, and project management, fostering a more individualized and efficient learning environment. The chapter further explores the potential impact of this course on preparing a future workforce for personalized medicine and the broader implications of AI integration into education and research in bioinformatics. It concludes with a discussion on prospects of AI in education, offering valuable insights into the potential of this technology in revolutionizing the learning process.

**Keywords:** bioinformatics, genomic medicine, AI in education, biopython, gpt-4.

## 1. INTRODUCTION

As personalized medicine gains prominence, bioinformatics, especially genomic medicine, is becoming increasingly vital in healthcare. Genomic medicine, leveraging individual genomic information for clinical care and policy implications, is revolutionizing disease understanding, diagnosis, and treatment. However, managing large-scale genomic data presents both opportunities and challenges, necessitating a proficient bioinformatics workforce. This chapter introduces an innovative approach to bioinformatics education for genomic medicine using BioPython, targeted at high school and early college students with no prior programming experience. The course, combining lectures, interactive exercises, and projects, actively engages students, providing direct experience with biological data. Notably, the course integrates GPT-4, a leading AI language model, into its design and teaching process, assisting with lesson planning, generating relevant examples, providing real-time code feedback, and fostering collaborative critical thinking and problem-solving, representing a pioneering effort in utilizing AI in education (Manolio et al., 2013; Green and Guyer, 2011).

## 2. THE NEED FOR BIOINFORMATICS EDUCATION

Bioinformatics, the application of computational tools to manage and analyze biological data, is a rapidly expanding field, driven by the surge in genomic data. It plays a crucial role in genomic medi-

cine, enabling us to make sense of vast amounts of data and uncover insights about health and disease. However, despite its importance, bioinformatics education is not a regular material in traditional high school and early college curricula. This gap in education leaves students unprepared for careers in this burgeoning field and creates a barrier to the wider adoption of genomic medicine (Mulder et al., 2018).

### **2.1. Current state of bioinformatics education**

The current state of bioinformatics education varies significantly across educational institutions and regions. In high schools and early college curricula around the world, bioinformatics is often absent or, at best, touched upon superficially. Traditional biology education tends to focus heavily on the fundamental concepts of life sciences, while computational aspects tend to appear in programming courses. This lack of integration leaves a substantial gap in students' understanding of how to apply computational tools to biological questions (Wilson et al., 2018).

The absence of practical, direct experience with bioinformatics tools and methods in the curriculum is another critical issue. Theoretical knowledge is essential, but without the opportunity to apply this knowledge, students may struggle to grasp the practical implications and applications of bioinformatics.

Finally, there is a lack of standardization and clear learning pathways in bioinformatics education. This can make it challenging for students to navigate their learning journey in this interdisciplinary field, which combines biology, computer science, mathematics, and statistics.

### **2.2. Importance of training in bioinformatics for genomic medicine**

Genomic medicine harnesses the power of individuals' genetic information to provide more precise and personalized healthcare. To fully utilize this genetic information, we need bioinformatics - to analyze, interpret, and make sense of the large-scale genomic data that underpin these medical applications (Heather and Chain, 2016).

Bioinformatics skills are key to understanding the genomic basis of disease. By analyzing genetic variants and gene expression data, we can identify potential biomarkers and therapeutic targets, leading to better diagnosis, prognosis, and treatment of diseases.

This knowledge aids in the interpretation and utilization of data from advanced technologies such as next-generation sequencing, which are now integral tools in medical research and increasingly in clinical practice.

Training in bioinformatics fosters critical thinking and critical thinking skills. It equips students to manage complex biological data and to extract meaningful insights from it, skills that are highly valuable in today's data-driven world.

With the growing demand for personalized medicine and data-driven healthcare, there is an increasing need for professionals with experience in bioinformatics. Therefore, integrating bioinformatics training into the education system is not only important, but imperative for the future of healthcare.

### **2.3. Projected future needs**

Expert analysts project that the demand for bioinformatics skills in genomic medicine will grow exponentially. With the continuing rapid advancements in sequencing technologies and the increasing adoption of personalized medicine, the generation of genomic data will accelerate. This deluge of data will necessitate a skilled workforce capable of handling, analyzing, and interpreting these data effectively (Karczewski et al., 2018).

There will be an increasing need for bioinformatics literacy across all levels of the healthcare profession, not just among specialists. Clinicians, nurses, and allied health professionals will need to understand the basics of bioinformatics to interpret genomic data and integrate these insights into patient care.

The integration of artificial intelligence and machine learning in healthcare will further amplify the need for bioinformatics skills. These technologies rely heavily on the processing and interpretation of large datasets – skills at the core of bioinformatics.

As research continues to uncover the genomic basis of disease, there will be an escalating demand for bioinformatics professionals in research and development, both in academia and the pharmaceutical industry.

### **3. PYTHON AND BIOPYTHON: TOOLS FOR BIOINFORMATICS**

As we explore the practical aspects of bioinformatics education, Python and BioPython emerge as indispensable tools for teaching and applying bioinformatics. In this section we describe the utility and accessibility of these tools, their role in bioinformatics, and why they are ideally suited for instruction in this interdisciplinary field (Cock et al., 2009; Model et al., 2005).

#### **3.1. Introduction to Python and BioPython**

Python is a high-level programming language known for its simplicity and readability, making it an excellent starting point for students new to coding. Its intuitive syntax, robust standard library, and active community support have made it one of the most popular languages for scientific computing and particularly bioinformatics.

BioPython is a set of freely available tools for biological computation written in Python. It is an open-source project that allows biologists and other scientists to use Python for bioinformatics and biological research. BioPython provides functionalities for a wide range of tasks such as reading and writing different sequence file formats, performing sequence alignment, manipulating DNA and protein sequences, and interfacing with common bioinformatics tools and databases.

#### **3.2. Importance of these tools in bioinformatics and genomic medicine**

Python and BioPython have become essential tools in bioinformatics and genomic medicine due to their versatility, ease of use, and the comprehensive set of functionalities they offer.

Python is a versatile programming language with a large standard library, making it suitable for a wide range of applications, from data manipulation and analysis to machine learning and web development. Its versatility enables researchers and practitioners to perform diverse tasks in bioinformatics and genomic medicine using a single programming language.

This programming language is known for its simplicity and readability, which lowers the barrier to entry for students and professionals from various backgrounds. This makes it an ideal choice for teaching bioinformatics, as students can quickly grasp programming concepts and focus on understanding the biological implications of their analyses.

BioPython, as a specialized library for biological computation, further enhances Python's capabilities by providing a suite of tools specifically tailored for bioinformatics tasks. These include functions for parsing and managing sequence data, performing sequence alignment, building phylogenetic trees, and accessing various biological databases. The availability of these tools in a single library streamlines the bioinformatics workflow and promotes efficient and reproducible research.

## 4. DESIGN OF AN INTRODUCTORY COURSE

This comprehensive five-module curriculum offers an excellent blueprint for teaching applied bioinformatics for genomic medicine. It presents a blend of theoretical knowledge and direct experience, ensuring students gain a firm grounding in the essential concepts of genomics and their applications in medicine, as well as practical skills in using Python and BioPython for bioinformatics tasks (Welch et al., 2014; Via et al., 2011).

The curriculum covers a wide range of topics, from the basics of genomic medicine and BioPython, to working with genomic databases, understanding genetic variations and diseases, analyzing gene expression, and applying machine learning in genomic medicine. Each module culminates in a lab project that allows students to apply the concepts they have learned in a practical setting.

### 4.1. Course objectives

The primary objective of this course is to equip students with a strong foundational understanding of bioinformatics, particularly as applied to genomic medicine. By introducing students to Python and BioPython, the course aims to provide them with practical computational skills that they can apply to the analysis and interpretation of biological and genomic data. A list of the key Biopython modules studied in the course appears in Table 1.

**Table 1.** Key Modules within the BioPython Library and Their Functions

| Module        | Description                                                                                                                                           |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Bio.Seq       | Provides a class for sequence objects, and tools for basic sequence manipulation.                                                                     |
| Bio.SeqIO     | Input/Output interface for reading and writing sequence file formats.                                                                                 |
| Bio.SeqRecord | Provides a class for Sequence Record objects, which contains a Seq object plus associated metadata.                                                   |
| Bio.AlignIO   | Input/Output interface for reading and writing sequence alignment file formats.                                                                       |
| Bio.Align     | Contains code to deal with multiple sequence alignments, including the construction of sequence alignments, such as the progressive alignment method. |
| Bio.Blast     | Tools for parsing BLAST output and running BLAST commands.                                                                                            |
| Bio.Entrez    | Interface for the NCBI Entrez database, allowing access to and parsing of returned data.                                                              |
| Bio.PDB       | Classes for the manipulation of Protein structure data.                                                                                               |
| Bio.Phylo     | Tools for working with phylogenetic trees, including visualizing, reading, and writing tree files.                                                    |
| Bio.SearchIO  | Input/Output interface for bioinformatics search tools like BLAST, BLAT, and HMMER.                                                                   |

The course aims to impart an understanding of the role and potential of genomics in medicine. This includes an exploration of how the use of genomic data informs diagnosis, treatment, and prevention strategies in a range of diseases, underpinning the growing field of personalized medicine.

The course also aspires to develop students' problem-solving and critical thinking skills. By engaging in interactive exercises and projects, students will learn to tackle real-world bioinformatics problems, fostering their ability to think critically and independently.

Finally, the course aims to demonstrate the potential of AI, specifically GPT-4, in education. It seeks to show how AI can enhance the learning experience by providing tailored educational content, offering immediate feedback, and facilitating student-educator interactions.



## **4.2. Course structure: lectures, exercises, and projects**

We designed this course with the objective of ensuring a comprehensive, yet engaging learning experience. It comprises a mix of lectures, interactive exercises, and practical projects, each of which plays a crucial role in the learning process.

Lectures form the backbone of the course, providing students with theoretical knowledge and conceptual understanding. They cover a range of topics, including the basics of Python and BioPython, the use of genomic databases, the role of genetic variations in diseases, the importance of gene expression, and the application of machine learning in genomic medicine. We designed the lectures to be interactive, encouraging students to ask questions, participate in discussions, and actively engage with the material.

The course integrates exercises throughout the lectures to allow students to apply what they have learned in a practical context. These exercises include programming tasks, sequence analyses, accessing and parsing data from genomic databases, and more. They provide students with immediate, firsthand experience with bioinformatics tools and techniques, and allow them to receive immediate feedback on their work.

Projects are a key component of the course, designed to consolidate learning and provide students with an opportunity to work on complex, real-world problems. These projects involve the analysis of biological and genomic data, encouraging students to work collaboratively, fostering teamwork and critical thinking skills.

## **4.3. Accessibility for students with no prior programming experience**

One of the fundamental design principles of this course is to make it accessible to students with no prior programming experience. The integration of Python, a language renowned for its readability and simplicity, aids in achieving this objective. Python's syntax closely resembles English, making it an excellent language for beginners.

The course uses the BioPython library extensively, which has the characteristic of being user-friendly. This allows students to focus on the biological problems they are trying to solve rather than grappling with the complexities of the code.

The use of GPT-4 further enhances the accessibility of the course. GPT-4 assists in creating tailored lesson plans, generating examples and exercises that cater to different skill levels. It also provides step-by-step explanations of programming concepts and immediate feedback on students' code.

The course also emphasizes the importance of a supportive learning environment. Students work together on exercises and projects, developing a collaborative learning culture where they can learn from each other.

## **5. COURSE CONTENT**

In this section we describe the content of the bioinformatics course. This comprehensive curriculum, spanning five modules, covers an array of topics from the basics of Python and BioPython to advanced concepts such as gene expression analysis and machine learning applications in genomic medicine (Lesk, 2013).

## 5.1. Biological sequence analysis: working with sequences, sequence alignments, phylogenetic analysis

Following the initial groundwork with Python and BioPython, students explore the intricacies of biological sequence analysis. This includes working with sequences, performing sequence alignments, and executing phylogenetic analyses. These are fundamental skills in bioinformatics and are critical for understanding the genomic data that underlies modern medicine.

Working with sequences is the bedrock of bioinformatics. Students learn how to import and manipulate biological sequences using BioPython's Seq objects. This includes understanding the properties of sequences, such as their length, composition, and the ability to slice and concatenate sequences.

Sequence alignment, another key topic, involves comparing two or more sequences to find regions of similarity. It is a fundamental technique in bioinformatics, used to infer functional, structural, or evolutionary relationships between sequences.

Students explore the concept of phylogenetics, the study of evolutionary relationships among species or groups of organisms. They learn how to construct and interpret phylogenetic trees, a tool that visually represents these relationships. We use BioPython's capabilities to manage phylogenetic trees, enabling students to perform tasks such as reading, writing, and visualizing phylogenetic trees.

Through these topics, students gain a practical understanding of sequence analysis. They learn not just the theoretical basis of these techniques, but also how to apply them to real biological data, providing them with the practical skills needed to work in bioinformatics and genomic medicine.

## 5.2. Genomic data exploration: databases, gene expression data, identification of disease-relevant mutations

Databases such as GenBank, Ensembl, the UCSC Genome Browser, dbSNP, and ClinVar, among others, are integral resources in genomic research. These repositories offer a wealth of genetic sequences, gene expression datasets, and variant information. The course teaches students how to access, parse, and analyze data from these databases using BioPython. Understanding these databases, their structure, and the information they contain is fundamental to bioinformatics.

Students analyze gene expression data, learning about its role in health and disease. They explore RNA-Seq technology, a widely used method for studying gene expression. By using Python and BioPython, students learn to access and analyze gene expression data from databases like GEO and SRA. This knowledge is crucial in understanding differential gene expression, which can provide insights into the molecular mechanisms of disease.

The course also covers the identification and analysis of genetic variations related to disease. Students learn about diverse types of genetic variations, such as Single Nucleotide Polymorphisms (SNPs) and Copy Number Variations (CNVs), and how they can contribute to disease. Using the BioPython toolkit, students learn to identify and analyze these variations, providing valuable skills for a future in personalized medicine. A list of the databases used in the course appears in Table 2.

**Table 2.** List of Online Genomic and Bioinformatics Databases Used in the Course with BioPython

| Name    | Description                                                                                                    | Web address                                                                               |
|---------|----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| GenBank | GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. | <a href="https://www.ncbi.nlm.nih.gov/genbank/">https://www.ncbi.nlm.nih.gov/genbank/</a> |

| Name                          | Description                                                                                                                                                                                                                                                                     | Web address                                                                               |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| Ensembl                       | Ensembl provides the genome resources for chordate genomes with a particular focus on human, mouse, zebrafish, and rat.                                                                                                                                                         | <a href="https://www.ensembl.org/">https://www.ensembl.org/</a>                           |
| UCSC Genome Browser           | The UCSC Genome Browser database is an interactive website offering access to genome sequence data from a variety of vertebrate and invertebrate species and major model organisms.                                                                                             | <a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a>                           |
| dbSNP                         | dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations. | <a href="https://www.ncbi.nlm.nih.gov/snp/">https://www.ncbi.nlm.nih.gov/snp/</a>         |
| ClinVar                       | ClinVar is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence.                                                                                                                                 | <a href="https://www.ncbi.nlm.nih.gov/clinvar/">https://www.ncbi.nlm.nih.gov/clinvar/</a> |
| COSMIC                        | The Catalogue of Somatic Mutations in Cancer (COSMIC) is a database of manually curated somatic mutation information relating to human cancers.                                                                                                                                 | <a href="https://cancer.sanger.ac.uk/cosmic">https://cancer.sanger.ac.uk/cosmic</a>       |
| GEO (Gene Expression Omnibus) | GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data collection.                                                                                                                                     | <a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>         |
| SRA (Sequence Read Archive)   | SRA is a public repository of sequencing data from all types of high-throughput sequencing platforms.                                                                                                                                                                           | <a href="https://www.ncbi.nlm.nih.gov/sra">https://www.ncbi.nlm.nih.gov/sra</a>           |
| PDB (Protein Data Bank)       | The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies.                                                                                                                                             | <a href="https://www.rcsb.org/">https://www.rcsb.org/</a>                                 |
| UniProt                       | The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data.                                                                                                                                                                  | <a href="https://www.uniprot.org/">https://www.uniprot.org/</a>                           |

### 5.3. Projects: description and objectives

An essential aspect of the course is the practical application of learned skills through projects. The design of the projects provides students with direct experience in bioinformatics, ensuring that they not only understand the concepts theoretically but can also apply them to real-world scenarios. Each project is associated with a specific module's content, reinforcing the knowledge and skills gained.

The five-module educational program consists of projects designed to enhance students' understanding of biological and genomic data analysis. In Module one, students work with biological sequences, performing sequence alignments using BioPython to infer evolutionary relationships. Module two engages them in retrieving and analyzing gene sequences from genomic databases to comprehend their role in disease. In Module three, the focus shifts to identification and analysis of genetic variations related to a specific disease using databases like dbSNP and ClinVar. Module four involves analyzing gene expression data, requiring students to identify differentially expressed genes and their potential role in disease conditions. Finally, in Module five, students apply machine learn-

ing to genomic data to predict disease outcomes or classify genetic variations, thereby exploring the convergence of AI and genomic medicine.

## **6. INNOVATIVE USE OF GPT-4 IN COURSE DEVELOPMENT AND INSTRUCTION**

In this section we describe the integration of GPT-4, a state-of-the-art AI language model, into course development and instruction. As we venture further into the age of artificial intelligence, we recognize the potential of AI in education. This section details the multiple roles GPT-4 plays in enhancing the course design, making learning more accessible, and fostering a conducive environment for active learning and critical thinking (Cheng et al., 2023; Lubiana et al., 2023; Gimpel et al., 2023).

### **6.1. Overview of GPT-4 capabilities**

Generative Pre-trained Transformer 4 (GPT-4) is a highly advanced AI language model developed by OpenAI. Machine learning techniques and transformer architecture power GPT-4, allowing it to generate human-like text with impressive coherency, accuracy, and context-sensitivity.

GPT-4 apart can provide immediate feedback on programming code, identifying errors, and suggesting improvements, making it a valuable assistant for teaching programming. In the course, GPT-4 facilitates discussions by providing relevant prompts and answering questions. It guides students through code debugging processes, helping them understand the logic and flow of algorithms. It can cater to a wide array of student interests and skill levels, making learning more personalized and effective.

### **6.2. Lesson planning**

One of the major roles that GPT-4 plays in the bioinformatics course is in lesson planning. Its ability to generate contextually relevant content allows it to create comprehensive lesson plans tailored to specific learning objectives. This includes not only the core content of the lessons but also the generation of relevant examples, exercises, and assessment materials.

The AI model's knowledge base and understanding of context allow it to provide a wide range of examples and exercises that cater to diverse student interests and skill levels. It creates both formative and summative assessments, providing critical feedback to both the students and the educators. This feedback can inform adjustments to the course material, ensuring that the pace and content are appropriate for the students' learning progress.

### **6.3. Teaching and explanations**

In the teaching process of the bioinformatics course, GPT-4 serves as a highly effective teaching assistant. GPT-4 provides detailed, step-by-step explanations of complex programming concepts, syntax, and best practices. It can elucidate intricate bioinformatics concepts and Python programming procedures in an accessible manner, breaking down the barriers that often hinder students without prior programming experience.

GPT-4 offers immediate feedback on students' code. By highlighting errors and suggesting improvements, it fosters a learning environment where students can learn from their mistakes in real-time. It also plays a significant role in facilitating class discussions. By providing relevant prompts and answering questions, it encourages critical thinking and problem-solving among students. It can offer insights, stimulate debate, and guide discussions in a way that reinforces the course material and helps students to connect theoretical knowledge with practical applications.

### 6.3.1. Debugging and algorithm understanding

One of the most significant challenges faced by students learning programming, particularly in a specialized field like bioinformatics, is debugging - the process of identifying and fixing errors in code. GPT-4 plays a pivotal role in this regard. Its advanced natural language processing capabilities enable it to read and understand code, making it capable of identifying errors and suggesting solutions. This not only helps students overcome hurdles in their coding tasks but also imparts valuable debugging skills that are essential in real-world bioinformatics work. An example of code analysis appears in Figure 1.

```
### Response with improving comments from GPT-4

In [ ]: # Import the necessary modules. Use their full names to make the code
# more readable and less prone to errors due to name conflicts.
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

# Define the model using a clear and descriptive variable name. Keep the model
# layers within the Sequential constructor for better readability and
# understanding of the model architecture.
model = Sequential([
    Dense(32, activation='relu', input_shape=(10,)),
    Dense(32, activation='relu'),
    Dense(1, activation='sigmoid'),
])

# Compile the model. The spaces around the equals sign are not necessary.
model.compile(optimizer='adam', loss='binary_crossentropy',
              metrics=['accuracy'])

# Print the model summary. A comment is not necessary here as the method
# name is self-explanatory.
model.summary()
```

Figure 1. Python code with improving comments from GPT-4 responding to a student's exercise

Beyond debugging, GPT-4 is instrumental in helping students understand the logic and flow of algorithms. Through step-by-step explanations, it guides students in deciphering how algorithms work, contributing to a deeper understanding of the principles of programming and data analysis.

In this process, students can enhance their coding proficiency, gain a solid understanding of algorithm design and optimization, and build essential skills for their future in bioinformatics and genomic medicine.

### 6.4. Project idea generation and project management

GPT-4 also helps students in forming creative ideas and in the tasks of project management. When it comes to formulating project ideas, GPT-4 proves to be an invaluable tool. It can generate a variety of project topics based on the course content, catering to diverse student interests and relevant real-world applications of bioinformatics. This allows students to engage with projects that not only consolidate their learning but also stimulate their curiosity and interest in the field.

In the area of project management, a crucial skill in any professional setting, it can help in organizing tasks, setting deadlines, and tracking progress, ensuring that group projects run smoothly and

efficiently. This not only helps to keep the projects on track but also imparts valuable project management skills to students.

In facilitating idea generation and project management, GPT-4 enhances the overall learning experience, making it more interactive and productive. It encourages student collaboration and fosters a sense of responsibility and professionalism, essential traits for their future careers in bioinformatics and genomic medicine. Figure 2 illustrates a project for the analysis of cancerous and healthy tissue samples.

### ### Download of analysis of gene expression data for cancerous and healthy tissue samples.

```
: # Load the dataset into a pandas DataFrame
df = pd.read_csv('gene_expression_data.csv')

# Display the first few rows of the DataFrame
print(df.head())

# Separate the cancerous and healthy samples
cancer_samples = df[df['condition'] == 'cancer']
healthy_samples = df[df['condition'] == 'healthy']

# Perform a t-test for each gene
p_values = stats.ttest_ind(cancer_samples.iloc[:, 2:],
                           healthy_samples.iloc[:, 2:], axis=0).pvalue

# Correct for multiple testing
adjusted_p_values = stats.multitest.multipletests(p_values, method='fdr_bh')[1]

# Add the adjusted p-values to the DataFrame
df['adjusted_p_value'] = adjusted_p_values

# Identify differentially expressed genes
differentially_expressed_genes = df[df['adjusted_p_value'] < 0.05]

# Print the differentially expressed genes
print(differentially_expressed_genes)
```

**Figure 2.** Python code for the analysis of cancerous and healthy tissue samples

## 7. BENEFITS AND CHALLENGES OF INTEGRATING GPT-4 IN EDUCATION

As we venture into the frontier of AI-integrated education, it is essential to evaluate both the potential benefits and challenges this novel approach brings. This section describes the specifics of integrating GPT-4, a state-of-the-art AI model, into the educational process. While the incorporation of AI in education promises a wide array of advantages such as personalized learning, improved accessibility, and enhanced engagement, it also poses certain challenges that educators and institutions must navigate. These challenges range from technical issues to ethical considerations and the evolving role of educators in an AI-enhanced teaching environment (Poldrack, Lu and Beguš, 2023; Liu et al., 2023).

### 7.1. Benefits: personalized learning, immediate feedback

The integration of GPT-4 into education offers measurable benefits that can revolutionize traditional learning methods. It allows for personalized learning experiences. GPT-4's ability to generate context-relevant examples and exercises facilitates a tailored approach that matches each student's pace, skill level, and interests. This personalized instruction has the potential to enhance understanding and retention of course material, making learning more effective.

GPT-4 provides immediate feedback on the student's work, a feature that is particularly beneficial in programming assignments. The model can quickly identify errors in code, suggest corrections, and provide explanations, thereby reducing the time it takes for students to troubleshoot and learn from their mistakes. This immediate feedback mechanism can significantly improve the learning curve and foster student confidence.

In addition, GPT-4 can enhance student engagement by creating interactive learning environments. By generating relevant prompts, answering questions, and facilitating discussions, GPT-4 encourages active participation and critical thinking. Furthermore, it can assist in project management, helping students organize tasks, set deadlines, and track progress, thereby fostering a sense of responsibility and professionalism. These benefits collectively represent a change in basic assumptions in education, opening new avenues for effective teaching and learning.

## **7.2. Challenges: technological hurdles, bias, and fairness**

Despite the transformative potential of integrating GPT-4 in education, we need to address measurable challenges. Technologically, ensuring all students have access to the necessary hardware and reliable internet is crucial, especially in remote learning scenarios. Similarly, the integration of AI models into existing educational platforms can present technical hurdles that require careful planning and resources to overcome.

Another significant challenge is the potential for algorithmic bias. AI models, including GPT-4, use for training vast amounts of data and can inadvertently learn and reproduce biases present in those data. Therefore, it is crucial to be aware of this issue and develop strategies to mitigate potential negative impacts on the learning process. This involves ensuring that the AI model treats all students fairly and does not favor certain types of questions, responses, or students over others.

Furthermore, there can be concerns about the over-reliance on AI for teaching, potentially diminishing the role of human educators. While GPT-4 can assist with lesson planning, generate examples, and provide feedback, it cannot replace the nuanced understanding, empathy, and guidance that human teachers bring to the learning environment. Balancing the use of AI with the irreplaceable human touch in education poses a considerable challenge.

## **7.3. Prospects of AI in education**

The future of AI in education holds immense promise, with potential advancements extending far beyond the current capabilities. The refinement of AI models like GPT-4 can further enhance the personalization of learning experiences, adapting instruction to cater to the individual needs, learning styles, and pace of each student. This could lead to improved academic outcomes and greater student engagement, as learning becomes more tailored and relevant.

AI's potential role in assessment and feedback could also be revolutionary. AI could help develop a more nuanced and comprehensive understanding of student progress, identifying not just what a student knows, but also how they learn best. This could inform more effective, tailored teaching strategies and interventions.

Also, AI could assist in democratizing education. By making quality education resources accessible to a larger audience, regardless of geographic location or socio-economic status, AI has the potential to bridge educational gaps.

Additionally, the integration of AI into education could also encourage the development of critical 21st-century skills among students, such as problem-solving, digital literacy, and adaptability. The future will see AI not just as a tool for education, but as a fundamental part of the educational landscape itself.

## **8. IMPACT AND FUTURE IMPLICATIONS**

This last section discusses the potential impact of integrating GPT-4 into bioinformatics education, specifically applied to genomic medicine, and explores the wider implications of this innovative approach for the future of education, research, and the genomics industry.

### **8.1. Potential impact on students and their preparedness for the workforce**

Incorporating GPT-4 into bioinformatics education can profoundly enhance students' skills and knowledge, preparing them for the ever-changing workforce. The course's practical nature allows students to transpose their learning into real-world situations, strengthening their problem-solving and critical thinking abilities. Combined with instant feedback from GPT-4, this experience can increase their confidence and proficiency in Python programming and bioinformatics, enhancing their competitiveness in the job market.

As personalized medicine becomes more prevalent, there will be an increasing demand for professionals who are experts at analyzing large-scale genomic data. By providing students with the necessary skills, this course plays a crucial role in shaping a future workforce ready to tackle the challenges of genomic medicine.

The use of advanced AI like GPT-4 in education could inspire students to explore the intersections of technology, biology, and medicine, potentially sparking innovative ideas and driving further advancements in these fields.

### **8.2. Implications for the future of bioinformatics and genomic medicine**

The integration of artificial intelligence in bioinformatics education signals a new frontier for both bioinformatics and genomic medicine. By equipping students with skills in Python programming, bioinformatics, and machine learning, this educational approach is creating a pipeline of talent ready to advance the field of genomic medicine.

With AI's ability to manage and analyze vast amounts of data, it can help in the development of new computational tools and methodologies, pushing the boundaries of what is currently achievable in genomics research.

This model of education could inspire the creation of similar courses in other areas of genomics and beyond, broadening the reach of AI-integrated education. This could lead to a more widespread understanding and application of genomics, potentially impacting areas such as disease diagnosis, treatment, and prevention, thereby transforming the future landscape of healthcare.

### **8.3. Future potential of AI and natural language processing in education and research**

Artificial Intelligence (AI) and Natural Language Processing (NLP), as demonstrated by GPT-4, hold considerable potential to transform education and research. The integration of AI in education can personalize learning, provide instant feedback, and engage students more effectively. It can supplement the role of educators, allowing them to focus on more complex instructional tasks, fostering critical thinking, and nurturing creativity.

In research, the capacity of AI and NLP to manage and interpret massive volumes of data can streamline data analysis, accelerate discoveries, and facilitate greater collaboration. For instance, AI can automate literature reviews, identify patterns and trends in data that humans might overlook, and even generate hypotheses for researchers to assess.



The potential of AI and NLP extends beyond individual courses or research projects. They are useful in the development of intelligent tutoring systems, create adaptive learning environments, and even predict learning outcomes.

## 9. CONCLUSION

This chapter has presented an innovative approach to teaching applied bioinformatics for genomic medicine, primarily targeting high school and early college students. By utilizing BioPython and integrating GPT-4, a state-of-the-art AI language model, the course offers a novel methodology that not only teaches bioinformatics but also highlights the potential of AI in education. The approach enables a more personalized, engaging, and interactive learning experience that accommodates diverse learning styles and paces.

In the era of personalized medicine, the importance of bioinformatics is growing exponentially, and the need for a skilled workforce capable of handling large-scale genomic data is paramount. This course aims to foster a new generation of bioinformaticians and genomic medicine professionals, equipping them with the necessary skills and knowledge.

The integration of AI in course development and teaching is still in its infancy, and this course represents an effort in that direction. While there are challenges to overcome, the benefits and potential are immense. As AI and NLP technologies continue to evolve, we can anticipate even more profound transformations in education and research, shaping the future of bioinformatics and genomic medicine.

## ACKNOWLEDGMENTS

The Basque Government funded partially this research with grant number IT1685-22.

## REFERENCES

- Cheng, K., Guo, Q., He, Y., Lu, Y., Gu, S., & Wu, H. (2023). Exploring the Potential of GPT-4 in Biomedical Engineering: The Dawn of a New Era. *Annals of Biomedical Engineering*, *51*, 1645-1653. <https://doi.org/10.1007/s10439-023-03221-1>.
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... & de Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422-1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N. & Vandrik, S. (2023). *Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education*. University of Hohenheim. <https://doi.org/10.13140/RG.2.2.20710.09287/2>.
- Green, E. D., & Guyer, M. S. (2011). Charting a course for genomic medicine from base pairs to bedside. *Nature*, *470*(7333), 204–213. <https://doi.org/10.1038/nature09764>
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. <https://doi.org/10.1016/j.ygeno.2015.11.003>
- Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease. *Nature Reviews Genetics*, *19*(5), 299-310. <https://doi.org/10.1038/nrg.2018.4>
- Lesk, A. M. (2013). *Introduction to Bioinformatics*. Oxford University Press. <https://global.oup.com/academic/product/introduction-to-bioinformatics-9780198794141>

- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., & Ge, B. (2023). Summary of chatgpt/gpt-4 research and perspective towards the future of large language models. *Cornell University, arXiv:2304.01852 (cs)*. <https://doi.org/10.48550/arXiv.2304.01852>.
- Lubiana, T., Lopes, R., Medeiros, P., Silva, J. C., Goncalves, A. N. A., Maracaja-Coutinho, V., & Nakaya, H. I. (2023). Ten Quick Tips for Harnessing the Power of ChatGPT/GPT-4 in Computational Biology. *Cornell University, arXiv:2303.16429 (cs)*. <https://doi.org/10.48550/arXiv.2303.16429>.
- Manolio, T. A., Chisholm, R. L., Ozenberger, B., Roden, D. M., Williams, M. S., Wilson, R., & Ginsburg, G. S. (2013). Implementing genomic medicine in the clinic: the future is here. *Genetics in Medicine, 15*(4), 258–267. <https://doi.org/10.1038/gim.2012.157>
- Model, F., Adorján, P., Olek, A., & Piepenbrock, C. (2005). Feature selection for DNA methylation based cancer classification. *Bioinformatics, 17*(suppl\_1), S157-S164. [https://doi.org/10.1093/bioinformatics/17.suppl\\_1.s157](https://doi.org/10.1093/bioinformatics/17.suppl_1.s157)
- Mulder, N., Schwartz, R., Brazas, M. D., Brooksbank, C., Gaeta, B., Morgan, S. L., Pauley, M. A., & Rosenwald, A. (2018). The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLOS Computational Biology, 14*(2), e1005772. <https://doi.org/10.1371/journal.pcbi.1005772>.
- Poldrack, R. A., Lu, T., & Beguš, G. (2023). AI-assisted coding: Experiments with GPT-4. *Cornell University, arXiv:2304.13187 (cs)*. <https://doi.org/10.48550/arXiv.2304.13187>.
- Via, A., De Las Rivas, J., Attwood, T. K., Landsman, D., Brazas, M. D., Leunissen, J. A., ... & Schneider, M. V. (2011). Ten simple rules for developing a short bioinformatics training course. *PLoS computational biology, 7*(10), e1002245. <https://doi.org/10.1371/journal.pcbi.100224>
- Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., & Schneider, M. V. (2014). Bioinformatics curriculum guidelines: toward a definition of core competencies. *PLoS computational biology, 10*(3), e1003496. <https://doi.org/10.1371/journal.pcbi.1003496>
- Wilson Sayres, M. A., Hauser, C., Sierk, M., Robic, S., Rosenwald, A. G., Smith, T. M., Triplett, E. W., Williams, J. J., Dinsdale, E., Morgan, W. R., Burnette, J. M., Donovan, S. S., Drew, J. C., Elgin, S. C. R., Fowlks, E., Galindo-Gonzalez, S., Goodman, A. L., Grandgenett, N. F., Goller, C. C., ... Pauley, M. A. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLOS ONE, 13*(6), e0196878. <https://doi.org/10.1371/journal.pone.0196878>