

Improving Clinical Bioinformatics Readiness among Healthcare Practitioners through AI-Enabled Education: A Pre–Post Study in Saudi Arabia

Mohamed Abouelhoda

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Anas Fallatah

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Mylynda Massart

University of Pittsburgh

Ahmad AbuSalah

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Ahmed Alfares

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Salah Baz

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Monther Alhamdoosh

`malhamdoosh@kfshrc.edu.sa`

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Workshop Team

King Faisal Specialist Hospital and Research Center (KFSH&RC)

Research Article

Keywords: Clinical Bioinformatics, Artificial Intelligence, Genomics, Continuous Medical Education, Healthcare Practitioners, Personalized Medicine

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10433190/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. Monther Alhamdoosh and Ahmed Abu AbuSalah declare that they were the scientific co-chairs of the training workshop being evaluated in this study, which represents a potential non-financial competing interest. The authors maintained objectivity in the data collection and analysis.

Improving Clinical Bioinformatics Readiness among Healthcare Practitioners through AI-Enabled Education: A Pre–Post Study in Saudi Arabia

Mohamed Abouelhoda^{1,*}, Anas Fallatah^{1,*}, Mylynda Massart², Ahmad AbuSalah³, Ahmed Alfares¹, Salah Baz¹, Monther Alhamdoosh^{1,*}, CBAIH2025 Workshop Team

¹ Genomic Medicine Center of Excellence, King Faisal Specialist Hospital and Research Center (KFSH&RC), Riyadh 12713, Saudi Arabia

² Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA 15260, United States

³ Digital Innovation Hub, King Faisal Specialist Hospital and Research Center (KFSH&RC), Riyadh 12713, Saudi Arabia

* Joint first co-authors

* Corresponding Author: malhamdoosh@kfshrc.edu.sa

Abstract

Background: Clinical Bioinformatics (CB) plays a pivotal role in the application of personalized medicine in clinical practice. Through genomics-based testing, CB contributes to various areas of healthcare by guiding diagnosis, treatment, and prevention. Despite the increasing interest in genomics education among healthcare professionals, there has been limited focus on providing targeted training in CB that addresses the needs of both patient-facing and non-patient-facing practitioners. This gap is largely due to the subject's complexity and the limited time available for practitioners' training. Furthermore, recent advancements in artificial intelligence (AI) have generated enthusiasm within the community regarding how AI can be responsibly leveraged to accelerate knowledge retrieval and strengthen CB skills for improved patient care.

Objective: In this paper, we demonstrate how recent advances in AI platforms can be utilized to deliver CB training to healthcare professionals. Our objectives are to: (i) enhance participants' understanding of the data and pipelines used in CB practice; (ii) strengthen their knowledge retrieval capabilities; and (iii) enable them to apply their limited computational skills to address complex data analytic tasks.

Methods: We developed a CB training workshop that integrates AI with traditional bioinformatics training. The workshop is designed to equip participants with skills that extend beyond basic genomics concepts and simple information retrieval. Our training prepares attendees for advanced analytics of complex -omic data and the self-development of data analytic software applications. In addition to technical content, the program addresses regulatory considerations, accreditation requirements, best practices, and frameworks for the responsible use of AI in healthcare.

Results: We organized a three-day educational event and conducted an evaluation study to assess the impact of our approach on CB and AI education for healthcare practitioners. Evaluation was based on mentors' feedback and on-site surveys measuring participants' self-confidence and attitudes before and after the workshop. Due to a technical glitch, pre- and post-workshop

surveys could not be linked at the individual level and were analyzed as independent samples. Despite this constraint, the educational intervention significantly improved participants' confidence across key competencies (Hedges' $g = 0.69$; $p < 0.01$), while maintaining strong positive baseline attitudes toward CB and AI.

Conclusions: Our workshop and evaluation study highlighted a growing interest among healthcare professionals in CB and AI, offering a scalable template for workforce development in precision medicine. The study demonstrated an increase in participants' self-confidence regarding their CB knowledge and their ability to leverage generative AI for script development. Additionally, the structured introduction of AI was effective in clarifying practical use cases while underscoring a "human-in-the-loop" paradigm for critical verification and responsible implementation in clinical practice.

Keywords: Clinical Bioinformatics; Artificial Intelligence; Genomics; Continuous Medical Education; Healthcare Practitioners; Personalized Medicine.

Introduction

Clinical and translational bioinformatics are relatively recent terms that generally refer to the application of bioinformatic methods in medicine and biomedical research. Translational bioinformatics focuses on research activities such as elucidating genotype-phenotype associations, understanding the molecular basis of disease progression, identifying drug targets, and developing personalized genomic-based therapies [1, 2]. Clinical bioinformatics is more concerned with routine genomic testing for clinical practice, and it is now expanding across a wide variety of clinical applications [3, 4, 5]. In medical genetics and many other medical specialties, Clinical Bioinformatics (CB) workflows identify rare variants associated with disease pathogenesis by analyzing genomic data. In oncology, CB workflows identify driver mutations and other genomic alterations related to disease carcinogenesis and report their effects on treatment response. For general medicine clinics, CB workflows for pharmacogenomic testing analyze genomic data to identify relevant genotypes that can guide medication and dose selection. For infectious diseases, CB workflows assist in detecting pathogens and identify antimicrobial resistance factors.

The expanding application of genomics in healthcare can be mainly attributed to Next Generation Sequencing (NGS) technologies, which have revolutionized translational research and clinical practice. For translational research, they provided valuable tools to decipher the multi-omic layers of biology, including the genome, transcriptome, proteome, and, indirectly, the metabolome. The resulting data provides insights into the disease pathogenesis and prognosis, supporting drug target discoveries and treatment optimization. For clinical practice, NGS technologies enabled accurate, efficient, and affordable routine clinical genomics testing to inform disease diagnosis and guide treatment pathways.

The launch of large-scale medical genome projects in many countries has marked the start of a new era in medicine [6, 7, 8, 9]. The primary objectives of these projects were to: (i) identify genomic variations responsible for disease, (ii) support treatment and drug discovery, and (iii) promote the adoption of personalized medicine. Notable examples of these projects include the 100,000 Genome Project in the UK, UK Biobank Project, the All of Us Project in the USA, the Saudi Human Genome Program, and the Qatari Genome Project in the Middle East. These projects emphasized the necessity to incorporate genomic knowledge and education into the healthcare system. As a result, efforts have been intensively applied to build genomic competencies among healthcare professionals [5, 6, 7]. Many academic programs, courses, and professional training modules have been developed. Many of these endeavors have been summarized and reviewed [10, 11, 12, 13, 14, 15]. These reviews detail the educational needs, content offered, and types of events (educational interventions) for participants from a wide range of non-genetic background

professionals across different countries. They also discussed methods for evaluating events, assessing the outcomes, and assessing the impact on patient care, in the short or long term.

Despite increasing interest in offering genomics education to healthcare professionals, there is limited effort focused on providing similar activities in Clinical Bioinformatics (CB). This is despite its critical role in modern genomics, amid increasing data volume and analysis complexity. The earliest training courses to address CB training for healthcare professionals were summarized in the following papers [16, 17, 18, 19, 20], which summarize the needs and concerns of healthcare professionals to better understand several key elements of the bioinformatics pipelines, including data formats, analysis results, and quality control (QC) parameters. For patient-facing clinicians, they also wanted to learn how to access relevant genomic information and how to interpret the results and make informed clinical decisions.

Educational Needs in Clinical Bioinformatics

As indicated in [16, 17, 18, 19, 20], the target audience is heterogeneous, and this requires the creation of tailored educational content that can attract different learner groups. In general, healthcare professionals can be categorized into two major groups:

1. *Patient-Facing Professionals*: These include clinicians, genetic counsellors, and nurses. The educational needs of this group are mainly focused on understanding basic terminologies and clinical genomic reports to extract clinical insights.
2. *Non-Patient-Facing Professionals*: These include, for example, clinical genetic scientists, bioinformaticians, technologists, and molecular pathologists. The educational needs of this group are mainly focused on understanding what happens behind the scenes in bioinformatics workflows and the regulatory compliance requirements related to it.

When designing a syllabus to cover these needs, there would be many educational topics and modules. Using traditional approaches, this would require significant time and multiple sessions. This would not be affordable for Continuous Professional Development (CPD), where professionals have limited time available for learning and development. Fortunately, the recent advances in Artificial Intelligence (AI) could facilitate the teaching of complex topics in a reasonable time, and in addition, improve computing skills and provide a means for continuous education after educational events.

The recent advances in Generative AI (GenAI) have become very evident to the average human through several renowned and easily accessible applications, including ChatGPT, Gemini, and Manus, among others. These systems are based on Large Language Models and are able to understand human questions to find answers in huge text databases [21, 22, 23, 24, 25, 26]. They also have content generation capabilities that can produce human-readable text, images or even executable computer source code in a very short amount of time. These GenAI systems have

further evolved to solve more complicated tasks by invoking functionalities from other systems to solve tasks of multimodal nature (e.g., solving a math problem, recognizing patterns, generating images.) [27, 28, 29]. Recently, GenAI platforms have expanded to provision “AI Agents” that can utilize chain of thought (COT) techniques to solve complex real-life tasks (e.g., booking tickets, finding restaurants, and planning trips) on top of the knowledge- and cognition-related tasks [30, 31, 32].

Impact of Our Research

Our study shows an analysis of survey results demonstrating how our educational intervention, based on the use of AI, could increase the confidence of healthcare professionals to understand bioinformatics workflows and results, analyze their own data, and develop their own software applications, using different levels of AI assistance; ranging from direct questions to AI-bots to the level of using GenAI for code generation. We also show how our course could demystify and remove ambiguities about the use of AI in healthcare in general and Clinical Bioinformatics to advance its application in patient care.

Methods

Study's Intervention Design

In this paper, we present an AI-based approach to deliver Clinical Bioinformatics training to healthcare professionals. To our knowledge, our approach has the following novel aspects:

- It is the first event to use AI as part of Clinical Bioinformatics training, supporting participants at levels beyond information retrieval to cover analytics and the self-development of tools.
- It is the first course to cover topics beyond computational and data techniques. It covers regulations, accreditation requirements, best practices and the frameworks for responsible AI.

We implemented our approach during a three-day event from December 17-20, 2025. The event was titled “Clinical Bioinformatics and Artificial Intelligence in Healthcare (CBAIH)”. It was offered as part of the Continuing Professional Development (CPD) training of King Faisal Specialist Hospital and Research Center (KFSH&RC), Riyadh, KSA. KFSH&RC is the largest tertiary care hospital in Saudi Arabia and the leading regional institute in implementing genomic medicine with a dedicated Centre of Excellence (CoE) for Genomic Medicine (GMCoE). Genomic testing at KFSH&RC is part of its daily clinical activities, with full integration of services and results into the patient’s electronic medical record [34]. Further, a key component of the KFSH&RC applied AI ecosystem is building a digital culture that educates staff on the sustainable and responsible use of AI capabilities for the purpose of advancing care, streamlining operations, and improving user experience. Despite this strong institutional commitment to genomics, bioinformatics and AI, a persistent challenge remains: many healthcare practitioners appreciate the value of Clinical Bioinformatics and AI, but they have not received a structured introduction or training in these areas. Moreover, many healthcare providers endorse the value of AI and genomics conceptually but lack practical confidence in applying these technologies. This gap between enthusiasm and readiness will delay the expansion of personalized medicine across clinical care.

The event was open to KFSH&RC employees and external healthcare professionals. The event included a variety of delivery and hands on teaching methods as recommended in the literature [16, 17]: face-to-face sessions mixed with minimized e-learning modules; brief, interactive, and engaging talks and sessions; maximized slots for hands-on sessions; group projects addressing general and individual attendee problems; as well as panel discussions with prominent experts. The event was accredited by the Saudi Commission for Health Specialties (SCFHS) and received Continuing Medical Education (CME) accreditation for 18 hours.

We organized a three-day educational event and conducted a pre–post educational evaluation study to assess the impact of a structured workshop on Clinical Bioinformatics and artificial

intelligence (AI) for healthcare practitioners (**Figure 1**). Below, we describe the learning objectives, the training form, the content of the topics, the Clinical Bioinformatics and the AI use cases, and the method for evaluating the event.

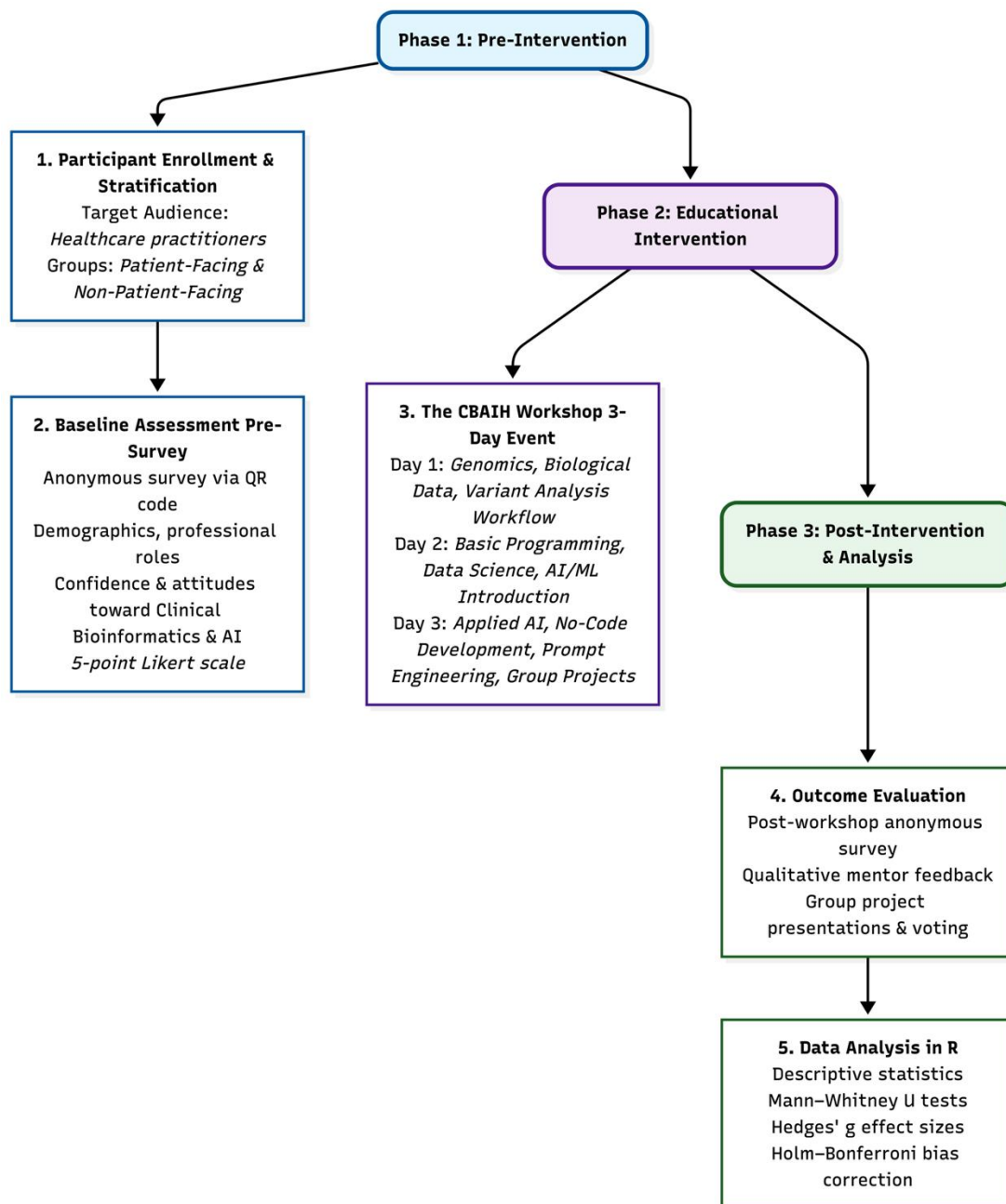


Figure 1: Overall study design and methodology

Learning Objectives

The learning objectives of the event were identified based on the literature [16, 17, 18, 19], and [20] which summarized the major needs and concerns of healthcare professionals. Moreover,

more targeted objectives were set to address the specific needs of our colleagues in KFSH&RC, particularly those using the in-house pipelines or receiving genetic reports. The final list of learning objectives is summarized in **Table 1** and mapped to Bloom's revised taxonomy [35], which organizes cognitive processes into six hierarchical levels: *remember, understand, apply, analyze, evaluate, and create*. The learning objectives were customized to the two major groups of audiences: 1) *Patient-Facing Professionals*, which included clinicians, genetic counsellors, and nurses, and 2) *Non-Patient-Facing Professionals*, which included, for example, clinical genetic scientists, bioinformaticians and molecular pathologists.

Table 1: Learning objectives of the workshop mapped to Bloom's revised taxonomy.

Target Group	Learning Objective	Bloom's Taxonomy Level
Patient-Facing	Understand the basics of genetics, genomics, as well as the genomic basis of diseases.	Remember, Understand
	Understand the principles of next-generation sequencing and the different genomic testing modalities.	Remember, Understand
	Understand genomic data types and Clinical Bioinformatics workflows.	Remember, Understand
	Identify biomedical knowledge by using genomic and clinical databases.	Remember, Understand
	Interpret genetic variation in a clinical context while understanding the limitations of bioinformatic analysis.	Remember, Understand, Apply, Analyze, Evaluate
	Understand ethical, legal and social implications of clinical use of genomic data.	Remember
Non-Patient-Facing	Understand basics of genetics, genomics, genomic basis of diseases, and DNA sequencing technologies.	Remember, Understand, Apply, Analyze
	Identify, compare, and analyze biomedical knowledge by comparing local and public genomics databases.	Remember, Understand, Apply
	Understand the details of the primary (base calling) and secondary (alignment and variant calling) bioinformatics workflows and its QC.	Remember, Understand, Apply, Analyze
	Analyze, annotate, and visualize the variants and identify the disease-causing mutations by performing filtration and prioritization.	Remember, Understand, Apply, Analyze
	Generate reports using standard tools or even customize reports by configuration or using self-developed AI tools.	Remember, Understand, Apply, Analyze, Create
	Apply data science skills to deal with big data files to extract clinical insights.	Remember, Understand, Apply, Analyze
	Develop and execute programming tools to analyze and interrogate data.	Remember, Understand, Apply, Analyze, Create
	Understand ethical, legal and social implications of clinical use of genomic data.	Remember, Understand, Apply

Training Form & Event Agenda

The training event included lecture-style talks, hands-on sessions, and panel discussions with prominent experts. The **INFO BOX 1** shows the event's agenda and scientific content. The talks on the first day were streamed online to enable wider participation from KFSH&RC. All hands-on projects were on-site, with face-to-face mentoring by professional clinical bioinformaticians. The time allocated to engaging discussions, hands-on activities, and projects was maximized, accounting for about 35% of the event.

The keynote talks introduced the fields of bioinformatics, Clinical Bioinformatics, multi-omics, and artificial intelligence, delivered by renowned scientists and experts in the field. The other topics and hands-on activities were introduced by senior domain experts and members of the KFSH&RC's Computational Sciences Department and Digital Innovation Hub.

INFO BOX 1

PART I

- **Introductory Topics**
 - Basics of Molecular Biology
 - Overview of Bioinformatics problems
 - Overview of Clinical Bioinformatics tasks in healthcare sector
- **Biological Data and Knowledge**
 - Basic Bioinformatics Databases (NCBI)
 - Medical oriented Databases (ClinVar, OMIM, etc.)
 - Retrieval of information
- **Clinical Bioinformatics in Action: Variant Analysis Workflow**
 - NGS technology
 - Variant Analysis Workflow [Primary, Secondary, Tertiary]
 - Variant Interpretation, Reporting, and Integration in Electronic Medical Record (EMR)
- **Beyond the computation and analysis**
 - QC of the results
 - Interpretation & Reporting of selected use cases
 - Accreditation and Regulation (CAP)
- **Panel Discussion**
 - Experts to motivate and guide
 - Clinical Side Story and Challenges (Head of Oncology, Head of Medical Genetics)
 - Future perspectives in the field

PART II

- **Responsible AI**
 - Overview of AI applications in healthcare
 - Responsible AI

- Regulations and Governance
- **Computing & Programming**
 - Introduction to Computing and Programming
 - Programming using a scripting Language: Python
 - Hands-on exercises
- **Data Science Programming Primer**
 - Basics of data types, representation, and visualization
 - Use of Python Data Science Capabilities
 - Hands-on & Exercises
- **Introduction to Machine Learning & AI**
 - Machine Learning Algorithms & Artificial Neural Network
 - Large Language Models
 - Hands-on & Exercises
- **AI in actions**
 - Overview of major AI platforms
 - Prompt Engineering
 - Generative AI
 - Large Language Models
 - RAG
 - Hands-on & Exercises

PART III

- The concept of No-Code Development
- Prompt Engineering to generate client side (Desktop) applications.
- Use of development frameworks to develop more complex applications: Gemini, ChatGPT, and Replit.
- Project: Variant Prioritization

The Main Use Cases: The Variant Analysis Workflow

We focused on variant analysis workflow which is the most widely used analytical workflow in Clinical Bioinformatics (**Figure 2**). The workflow is composed of the following steps:

- **Primary Analysis:** This includes the base calling steps to transform the physical signals (produced by the sequencing machine) into base calls in the form of a sequence of letters representing the sequence of DNA nucleotides in a fragment of DNA. The output of this step is a set of strings (also referred to as reads) representing each DNA fragment.

- **Secondary Analysis:** This includes the alignment of the sequencing reads to a reference human genome. This step is followed by variant calling to identify the variations in the input genome as compared to the reference genome.
- **Tertiary Analysis:** This step includes the annotation of the variant with all possible information. This information includes physical information about the location in the genome, in the related genes, the transcript, and the respective protein. It also includes information on the variant's frequency in population databases, its known or predicted effect on protein structure, and its known relationship to disease.
- **Filtration & Prioritization:** This step involves the exclusion of variants not related to the phenotype of interest and prioritizing the remaining variants according to the given clinical features of the patients.
- **Reporting:** This step involves generating the report and adding insights and recommendations on the identified variants.
- **Electronic Medical Record (EMR) Integration:** This step involves integrating the reports into the electronic medical record system to become part of the patient's medical record.

On the first day, the variant analysis workflow was presented through a series of lectures and talks, during which users participated in discussions and mini quizzes. The content of these lectures is shown in **INFO BOX 1**, which shows a series of lectures walking through this workflow. A dedicated lecture was held to discuss public biological databases, particularly ClinVar, to find information on variants and their ACMG classifications.

On the second day, the basic programming and AI methods were introduced so that the user can apply these skills to handle big data using Python, retrieve knowledge, and use basic AI services (ChatGPT and Gemini). The examples and discussion on the second day focused on genome sequencing tests and the output of the variant analysis workflow, showing how scientists go

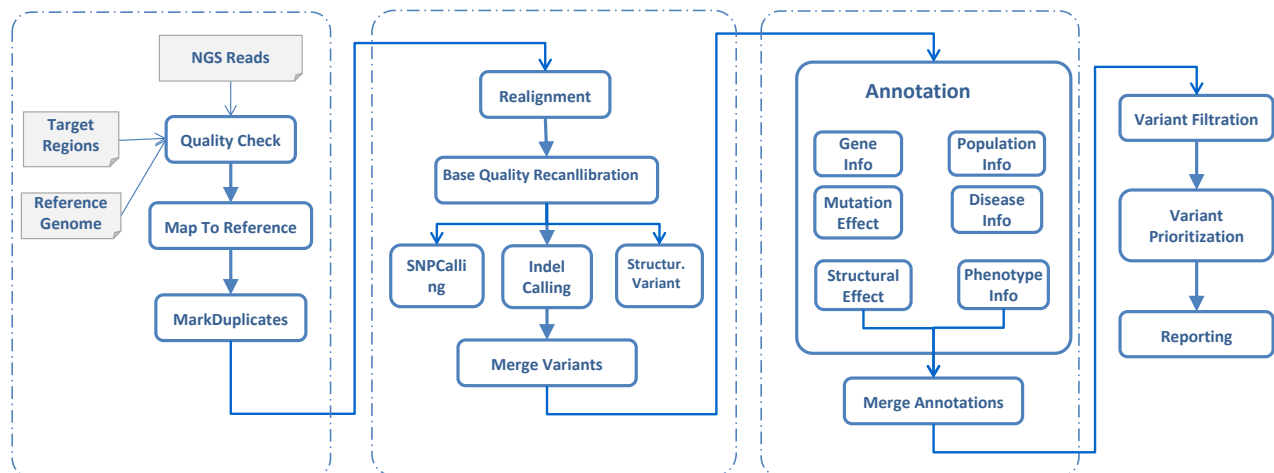


Figure 2: Variant Analysis Workflow. Primary analysis includes base calling and alignment. Secondary analysis includes variant calling. Tertiary analysis includes variant annotation. Filtration and prioritization are important to help the scientist spot the disease related mutations. Reporting is a critical step in the medical domain which includes the results to guide the diagnosis, support physician decisions and inform the patients. In KFSH&RC, the request of the test and the report submission are integrated with the hospital information system and accessible from in the patient electronic record.

beyond Excel and use scripting languages to search, sort, and filter their files. **INFO BOX 2** shows examples. At the end of the second day, the examples were about using AI services (ChatGPT and Gemini) to retrieve information about the variants and compare this to the public OMIM and ClinVar databases. Some tips about writing “Prompts” for these AI systems were addressed in an engaging discussion with live demos.

INFO BOX 2

BASIC PROGRAMMING EXAMPLES

Example 1: Use Python and write a program to open a list of tests conducted in the lab and identify whole-exome sequencing (WES) tests.

```

import csv
file = open('example.csv', mode='r', newline='')
reader = csv.reader(file)

for row in reader:
    if ("WES" in row[6]):
        print(row)
  
```

Example 2: Use Python and write a program to open an annotated VCF file in CSV format and filter out all non-coding variants.

Solution:

```
Import CSV
file = open('example.csv', mode='r', newline='')
reader = csv.DictReader(file)

for row in reader:
    if ( "exon" in row['type'].lower() ):
        print(row)
```

NOTE: In Example 2, the code was optimized to read the columns by header title using “csvDictReader” function, and also we made the check for the keyword exon case-insensitive using “lower()” function on the stored field.

ADVANCED PROGRAMMING EXAMPLE

Example 1: Use the Pandas library of Python to write a script to open a list of tests conducted in the lab and identify.

```
import pandas as pd
df = pd.read_csv("example.tsv", sep="\t")
# Filtering the names of the tests only unique
df_unique_names = pd.DataFrame(df['Test'].unique())

# Print the names
display(df_unique_names)

# Filter on the column
filtered_df = df_vcf[df_vcf["Type"] == "exon"]

# Print the Dataframe
display(filtered_df)
```

Example 2: Use the Pandas library of Python to write a script to open an annotated VCF file in CSV format and filter out all non-coding variants.

```
import pandas as pd

# Read the VCF file
df_vcf = pd.read_csv("example.vcf", comment="#")

# Filter on the column
filtered_df = df_vcf[df_vcf["Type"] == "exon"]
```

```
# Print the Dataframe
display(filtered_df)
```

Example 3: Use the Pandas library of Python to write a script to count how many WES tests was performed between June and July 2024.

```
# Condition that all Tests should be "WES" only
condition = df['Test'] == 'WES'
```

```
# Which months need to be added together
columns_to_sum = ['Jun', 'Jul']
```

```
# Total sum of the defined months
conditional_sum = df.loc[condition, columns_to_sum].sum(axis=1)
```

```
# Print
print(conditional_sum)
```

NOTE: In Example 1&2, the trainee was introduced to the features of Pandas. In Example 3, the trainee can conduct more complex query on the data, which would be more complicated to program without Pandas.

AI Use Cases

On the third day, the participants were invited to develop applications for the filtration and prioritization of variants using a no-code development strategy based solely on public AI platforms. The first session of that day demonstrated how this can work using prompt engineering to generate client-side programs. In subsequent project sessions through the end of the day, the trainees were required to develop more complex systems with advanced features and, if possible, generate a deployable server-side version. **INFO BOX 3** shows the three stages of these activities. For the projects, the trainees were asked to form groups of up to 4 participants and either select projects from a suggested list or identify related problems of their own. Each project was asked to present its output in the form of a short presentation and a quick demo.

INFO BOX 3

TASK 1: PROMPT ENGINEERING

Provide ChatGPT with a set of variants to generate professional-like scientific text (narrative) about the variant, effect, and medical related info.

Engineered Prompt to ChatGPT to generate the text

You are a clinical genomics assistant. Write a narrative for the following set of variants defined with the following fields: Gene, Variant, priority (High/Med/Low), rationale, flags (array), next_steps (array), Phenotype: {{PHENOTYPE_TEXT}} Variants: {{VARIANT_ROWS_JSON}}.

The narrative should include professional-like text describing the variant based on the fields uploaded to you. Do not hallucinate. If the field is empty, then write “unknown”..

TASK 2: CLIENT-SIDE SCRIPT

Build a browser-only vanilla JS web app (no frameworks) for this challenge: Rare Disease Variant Prioritization.

Application Requirement:

1. Upload variants.tsv (columns: Gene, Variant, Consequence, ClinVar, gnomAD_AF, Phenotype_Match)
2. Paste phenotype text
3. Explore ranked variants (filters + detail panel)
4. Export top hits + summary

Engineered Prompt to ChatGPT to generate the Application

You are a clinical genomics assistant. Do not hallucinate. If evidence is missing, write “unknown”.

Using the phenotype and the variant table, return ONLY valid JSON array with fields: rank, Gene, Variant, priority (High/Med/Low), rationale, flags (array), next_steps (array).

Phenotype: {{PHENOTYPE_TEXT}}

Variants: {{VARIANT_ROWS_JSON}}

Ranking guidance: prioritize lower gnomAD_AF, ClinVar Pathogenic/LP, LoF consequences, phenotype match.

UI components (must have)

- TSV uploader + table preview
- AF threshold slider
- Filters: ClinVar, Consequence, Phenotype_Match
- “Top N” selector (5/10/20)
- Ranked table + “Variant Card” detail panel

Deliver exactly 3 files: index.html, styles.css, app.js.

Must include: File upload (CSV/TSV) using FileReader

Table preview with search + sortable columns Filters relevant to the challenge

Row click -> detail panel

Export All results to JSON + CSV

Provide minimal clean styling, responsive layout, and dummy data fallback.

IMPORTANT: the app must run by opening index.html.

NOTE: You can use this list and use the prompt in TASK 1 to generate narrative about the highest priority variants. One can use the API feature from the APP to invoke ChatGPT on the data computed by the APP and using the prompt in task 1.

TASK 3: ADVANCED APP (PROJECT TASK)

Develop a server-side application where the user can submit their files and get the results.
Hint: Use ChatGPT plus Replit Node proxy version. The participants are free to use other AI-based platforms if they wish.

Method Evaluation

We used two methods to assess the outcome of the event: the first was qualitative, involving collecting feedback from the tutors/mentors who accompanied the trainees during the event exercises and projects, and from the attendees who presented their projects at the end of the event and voted on the best ones. The second was quantitative, based on a structured survey measuring levels of self-confidence and attitudes [33] toward Clinical Bioinformatics and AI in healthcare.

Survey Design

We used anonymous non-matched surveys distributed before and after the workshop to evaluate changes in self-reported confidence and attitudes [33] regarding the event's topics. The survey questions covered three components:

1. Demographic and professional background: In this section, attendees reported their academic backgrounds and their primary professional roles in the healthcare system. A major point to determine from this inquiry is whether they are patient-facing or not.
2. Confidence Items: Questions with multiple-choice answers to report self-confidence in Clinical Bioinformatics and AI competencies (e.g., understanding workflows, variant interpretation, AI/ML concepts, reproducibility). Responses were recorded on a 5-point Likert scale ranging from Not Confident at All (1) to Extremely Confident (5).
3. Attitude (agreement) Items: Questions with multiple-choice answers assessed perceptions of the importance, relevance, and role of Clinical Bioinformatics and AI in healthcare and precision medicine. Responses were recorded on a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5).

The survey questions are in **Supplementary File 1**. The survey was distributed to users via a QR code and completed online. The individual responses were anonymized and not linkable across time points. No identifiable personal, clinical, or patient-level data were collected. The study was conducted as an educational quality-improvement initiative and is aligned with institutional policies on educational evaluation. It was deemed not to require ethical approval.

Data Analysis

Descriptive statistics were used to summarize participant characteristics and survey responses, including frequencies, percentages, means, standard deviations, medians, and interquartile ranges. Non-parametric Mann–Whitney U tests were used to compare pre- and post-workshop responses. The Hedges’ g statistic was used to measure effect sizes, as it accounts for unequal sample sizes between groups. The Hedges’ g was used in this study to emphasize educational impact rather than statistical significance alone. Analyses were conducted at the individual components alone (demographic, confidence, and attitudes) and on the composite level (demographic, confidence and attitudes) as well to explore domain-specific trends. To avoid multiple-comparison problems and bias, the Holm–Bonferroni method was used to support transparent interpretation. All statistical analyses and figures were generated using R (version 4.5.2). Because individual responses were anonymized and not linkable across time points, pre- and post-workshop responses were analyzed as independent samples.

Results

Participant Characteristics

A total of 41 participants were registered for the event and completed the pre-event survey. Twenty-Two of them (54%) completed the post-event survey. The patient-facing group accounted for 24.4% of attendees, while the non-patient-facing group accounted for 75.6%. The non-patient-facing group included scientists/academics (51.2%) and bioinformatics or IT professionals (9.8%), technologists or specialists (7.3%), and trainees (4.9%). **Table 2** summarizes information about the participants.

Participants had different levels of field experience: 31.7% (13/41) had ≥ 8 years of experience after graduation, 29.3% (12/41) had 3–7 years of experience after graduation, and the remaining 39.0% (16/41) were fresh graduates and fellows.

Many participants reported little experience (18/41, 43.9%) or no experience (14/41, 34.1%) in bioinformatics or AI. Only 9.8% (4/41) reported a good amount of experience, 9.8% (4/41) reported some experience, and 2.4% (1/41) reported extensive expertise. Consistent with these findings, hands-on scientific computing usage was low at baseline. Most participants reported that they never used a command-line interface (24/41, 58.5%) or programming languages such as Python or R for data analysis (25/41, 61.0%). Fewer than 15% of participants reported frequent (often or always) use of either command-line environments or programming tools.

Collectively, these baseline characteristics indicated a cohort with strong clinical and scientific expertise but limited readiness in computational and AI, underscoring the need for accessible, implementation-focused education to increase uptake and use of Clinical Bioinformatics and AI capabilities across the healthcare workforce.

Table 2. Baseline participant characteristics and computing readiness (pre-workshop; N = 41).

Characteristic	Category	n (%)
Participant Group	Non–patient-facing	31 (75.6%)
	Patient-facing	10 (24.4%)
Primary Professional Role	Scientist/Academic	21 (51.2%)
	Clinician/Physician	10 (24.4%)
	Technologist/Specialist	4 (9.8%)
	Bioinformatics/IT	3 (7.3%)
	Student/Trainee	3 (7.3%)
Clinical Domain Expertise	Extensive experience (8+ years after graduation)	13 (31.7%)
	Little experience (1-2 years after graduation)	10 (24.4%)
	Some experience (3-4 years after graduation)	8 (19.5%)
	No experience (0 years after graduation)	6 (14.6%)

Table 2. Baseline participant characteristics and computing readiness (pre-workshop; N = 41).

Characteristic	Category	n (%)
	Good experience (5-7 years after graduation)	4 (9.8%)
Bioinformatics / AI Expertise	Little experience	18 (43.9%)
	No experience	14 (34.1%)
	Good experience	4 (9.8%)
	Some experience	4 (9.8%)
	Extensive experience	1 (2.4%)
Scientific Computing Usage — Command shell usage	Never	24 (58.5%)
	Rarely	10 (24.4%)
	Often	4 (9.8%)
	Sometimes	3 (7.3%)
Scientific Computing Usage — Programming usage	Never	25 (61.0%)
	Rarely	9 (22.0%)
	Sometimes	4 (9.8%)
	Often	2 (4.9%)
	Always	1 (2.4%)

Baseline Attitudes & Confidence Toward Clinical Bioinformatics and AI

Participants demonstrated strongly positive attitudes toward the role of Clinical Bioinformatics and artificial intelligence (AI) in healthcare and precision medicine. The mean agreement composite score was 4.05 (SD 1.00) on a 5-point Likert scale, indicating high endorsement of concepts such as the clinical value of genomics, AI-assisted decision-making, and the importance of bioinformatics capacity-building.

In contrast, baseline self-reported confidence in applying Clinical Bioinformatics and AI skills was considerably lower, with a mean confidence composite score of 2.04 (SD 0.99). This discrepancy highlights a gap between conceptual support for precision medicine and practical readiness to implement it.

Baseline attitudes were lower in patient-facing professionals (mean 3.80) than non-patient-facing professionals (mean 4.16).

Changes in Attitudes Toward Clinical Bioinformatics and AI Following the Workshop

Post-event attitudes toward Clinical Bioinformatics and AI remained highly positive. The composite score increased modestly from 4.05 to 4.27, representing a 0.22-point increase. This change did not reach statistical significance ($p = 0.35$, Hedges' $g = 0.23$), consistent with a ceiling effect given the already high baseline endorsement (**Figure 3**). The technical exposure discussed in the event did not make the participants feel that the topics are too complex and this did not reduce their enthusiasm. Therefore, the engagement with bioinformatics and AI concepts and

expertise was reinforced rather than undermined the participant's attitude. In fact, the largest gains in attitude were observed in items related to: Importance of AI/ML in Healthcare [Q18], Role of Collaboration in Precision Medicine [Q21], and Importance of Bioinformatics Best Practices [Q19].

Following the workshop, attitudes increased directionally in patient-facing participants (3.96 → 4.26) more than in non-patient-facing participants (4.07 → 4.27). Although these changes were not statistically significant, it shows some trend and it is consistent with high baseline endorsement (**Figure 3**). **Supplementary File 2** includes all the statistics results in tabular format.

Changes in Self-Reported Confidence Following the Workshop

The survey results following the educational intervention showed that the participants reported significant improvements in confidence across both Clinical Bioinformatics and AI competencies (**Figure 4**). The mean composite confidence increased from 2.04 (SD 1.18) pre-event to 2.71 (SD 1.11) post-event, representing a 0.67-point increase on a 5-point scale. This increase represented a medium-to-large effect size (Hedges' $g = 0.69$) and was statistically significant using a Mann–Whitney U test ($p = 0.006$).

The largest gains in confidence were observed in items related to: Knowledge of Variant Clinical Significance [Q11], Knowledge of Deep Learning and GenAI [Q14], Knowledge of Machine Learning [Q13], Knowledge of Bioinformatics Best Practices [Q15], Knowledge of Variant Calling Pipelines [Q10], and Knowledge of Biological Data Types [Q8]. These areas corresponded directly to the workshop's applied, implementation-focused curriculum.

Patient-facing participants showed a pronounced increase in AI-related confidence (1.7 → 2.6; Hedges' $g = 0.79$; $p = 0.161$), while non-patient-facing participants improved in both Clinical Bioinformatics–related confidence (2.10→2.78; $p = 0.0165$) and AI-related confidence (1.63→2.41; $p = 0.0131$) (**Figure 4**). **Supplementary File 2** includes all the statistics results in tabular format.

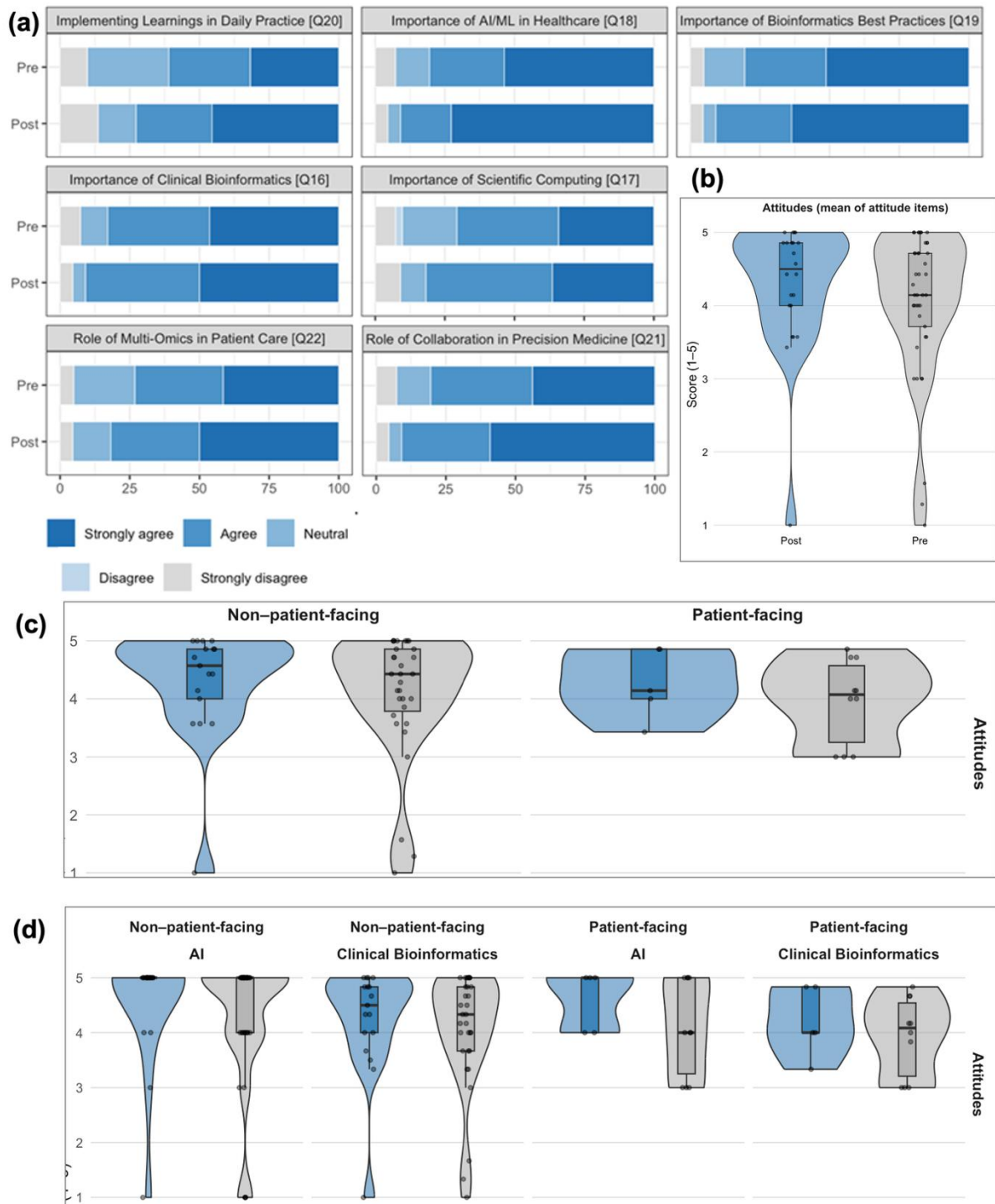


Figure 3: Self-reported attitude analysis. (a) Attitude and agreement pre and post the event (percent of responses per item). (b) Composite pre and post changes. (c) Attitude outcome stratified by patient and non-patient facing groups. (d) Further stratification by major topics (Clinical Bioinformatics and AI).

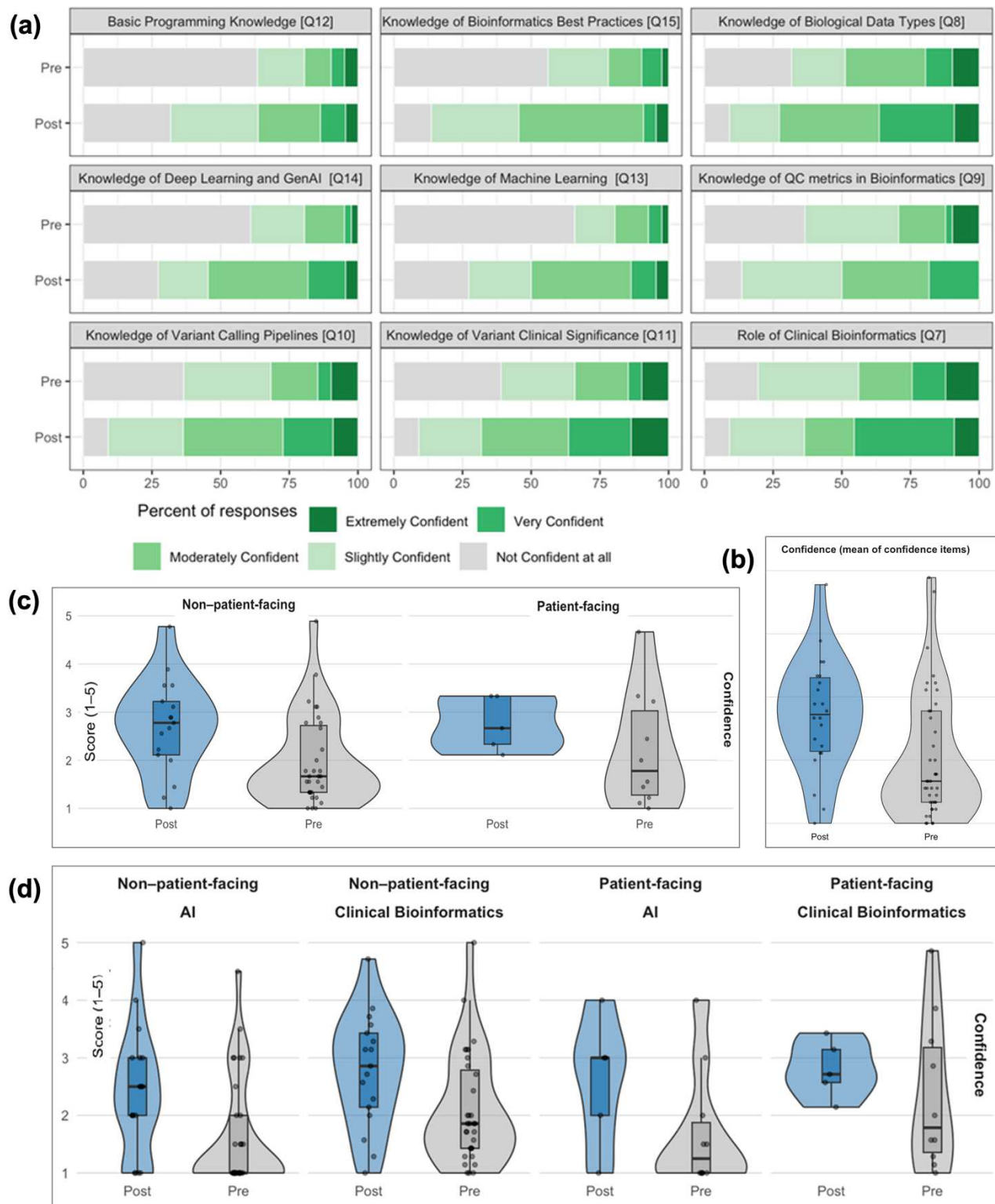


Figure 4: Self-reported confidence analysis. (a) Confidence outcome pre and post the event (percent of responses per item). (b) Composite pre and post changes. (c) Confidence outcome stratified by patient and non-patient facing groups. (d) Further stratification by major topics (Clinical Bioinformatics and AI).

Mentor feedback

Mentors with bioinformatics and computer science training background supervised the participants in conducting the workshop exercises and project. On the third day, we focused on the concept of GenAI to generate self-developed applications solely through prompt engineering, a concept known as No-Code development. As explained in the Method section, there were two major tasks: (i) develop a simple client-side application that can be accessed via a local web browser, where the user uploads their own data to filter the variants, and (ii) develop more advanced server-side application that can be deployed on a remote web server to be accessed by multiple users over the Internet.

Participants promptly grasped the concepts and developed the first application easily. **Figure 5** (Left) shows a screenshot of an example application. Some used Gemini and some used ChatGPT. Generating the narrative through the copy-and-paste option from the workshop learning materials to ChatGPT was practical for many of the groups. Automating this step via an API was somewhat complicated due to security keys, but many groups still finished the task on time, while others required intensive support from the mentors. We observed that some of the users did not edit the prompt but took a photo from the instructor board and submitted it as is to ChatGPT, which performed digital Optical Character Recognition (OCR), and transformed the image into text including the prompts, and generated the application source code.

The development of the server-side application was more challenging and required more time. Some of the groups could finish the task on time. Other groups could understand the basic principles but did not have time to finalize all features. Two groups produced output that could be deployed. One group deviated from the suggested procedure of using ChatGPT and Replit. Instead, they used Google AI Studio, which received prompts, generated the application source code, and performed deployment. **Figure 5** shows an example page of the application after filtering and reporting the variants using Replit and Google AI Studio.

Overall, participants demonstrated noticeable contributions in applying AI tools and began developing their own applications. While this training might have not enabled them to become professional software developers, it did allow them to independently create simple applications that handle tabular data, compare values, and prioritize genomic variants. The intervention significantly improved participants' confidence across key competencies, while maintaining strong positive attitudes toward Clinical Bioinformatics and AI.



Rare Disease Variant Prioritization

1. Input Data

Upload variants.tsv No file selected.

Patient Phenotype

Paste clinical notes here... (e.g., Hypotonia, developmental delay, seizures)

2. Filters & Analysis

Max gnomAD AF: 0.01

ClinVar

All

Top 5

Top N Results

Ranked Variants

Rank	Gene	Variant	ClinVar	AF	Priority
-	SCN1A	c.123G>A	Pathogenic	0.0001	N/A
-	BRCA2	c.456T>C	Likely Pathogenic	0.02	N/A
-	GJB2	c.35delG	Pathogenic	0.005	N/A

Variant Triaging Assistant

Rare Disease Variant Prioritization - Browser-only - No PHI

Demo / Workshop MVP

0) OpenAI API (Optional)

Warning: For workshops only. Do not ship production apps with API keys in the browser. Use a server-side proxy in real deployments. [contentReference\[paicr22\]\[index2\]](#)

OpenAI API Key Model **gpt-5-mini** Run with OpenAI API

Save Clear Pick any model you have access to in your OpenAI project.

Stored in local storage on this device only.

Success: AI results loaded from OpenAI API.

1) Load Data

Upload CSV/TSV Choose file No file chosen

Expected columns: Gene, Variant, Consequence, ClinVar, gnomAD_AF, Phenotype_Match

Load Dummy Data Clear Loaded 6 variant(s). Showing 6

Safety: Do not upload real patient data. This app is for training/demo only. Always validate with trusted sources.

2) Explore Variants

Search ClinVar Consequence Phenotype Match

All **All** **All**

Max gnomAD_AF

e.g., 0.001

Gene	Variant	Consequence	ClinVar	gnomAD_AF	Phenotype_Match
KMT3D	p.Arg543His	missense	VUS	0.00002	High
TTN	p.Trp23*	stop_gained	Pathogenic	0.00001	Low
SCN2A	p.Ala1207Trp	missense	Likely Pathogenic	0.00003	Medium
BRCA1	p.Gln776*	stop_gained	Pathogenic	0.000001	High
ANKRD11	p.Gln1010*	stop_gained	Likely Pathogenic	0.000005	High
MECP2	p.Trp138Met	missense	Pathogenic	0.0	Medium

Tip: Click a column header to sort. Click a row to view details.

6 variant(s) shown

BioVariant AI
Interpretation Suite

Unified GRCh38 genomic evidence and phenotype correlation.

EXECUTIVE FINDINGS

"Diagnostic confirmation of Retinal Arterial Macroaneurysm with Supravalvular Pulmonic Stenosis (IGFBP7). Cardiology evaluation recommended. Patient is a carrier for MPS VII (GUSB)."

Pathogenicity Profile

Pathogenic: 2
Likely Path.: 0
VUS: 1

Primary Findings (Causative)

TARGET GENE
POLR2B;IGFBP7

GRCh38 chr4:57031337
(p.Protein change unmapped)

ACMG EVIDENCE

PVS1 PS1 PP4

CLINICAL ACTION

Cardiology referral to evaluate for supravalvular pulmonic stenosis. Genetic counseling for autosomal recessive inheritance.

Figure 5: Example of data analytic applications developed by trainees. (a) Example a client-side application for prioritizing the variants. (b) Example of a server-side application developed using Replit to prioritize the variants, generate narrative, and export reports in PDF/HTML format. (c) Example server-side application developed using Google AI studio to prioritize the variants, generate narrative, and export reports in PDF/HTML format.

Discussion

This study demonstrates that a short, targeted 3-day educational intervention can meaningfully improve self-reported confidence in Clinical Bioinformatics and AI among a heterogeneous group of healthcare practitioners and researchers. The workshop successfully translated conceptual support into practical readiness, enabling participants to leverage Generative AI tools (such as ChatGPT and Gemini) for information retrieval, script generation, and workflow automation, including Variant Analysis. While participants entered the workshop with a high baseline endorsement of precision medicine, the training yielded the largest confidence gains in applied competencies—specifically variant interpretation, AI-assisted analysis, and workflow reproducibility—rather than abstract awareness.

The pronounced post-workshop confidence gains, particularly among non-patient-facing professionals, suggest that democratized, implementation-oriented education effectively bridges the gap between theoretical knowledge and practical execution. Notably, the integration of AI shifted the participants' cognitive load from mechanistic tasks, like manual script editing or data retrieval, to higher-order professional skills, including critical thinking, rigorous code testing, and clinical verification. Because Generative AI tools are prone to hallucinations, the workshop emphasized a "human-in-the-loop" paradigm. Participants were trained not to accept AI outputs blindly, but rather to treat AI as an accelerant while executing rigorous code execution testing and cross-referencing information against curated clinical data repositories. Furthermore, multi-disciplinary, collaborative groups performed better by sharing diverse individual experiences. This cross-pollination of ideas accelerated learning and highlighted how shared foundational training can unify clinicians, scientists, bioinformaticians and technologists who possess distinct professional needs.

These findings align with established medical education literature indicating that experiential, context-driven, and hands-on learning is uniquely effective for complex digital health skills. The success of incorporating real-world scenarios and data-driven bioinformatics problems replicates prior evidence that active learning sustains high participant focus and engagement. By focusing the Variant Analysis workflow on standard clinical frameworks—such as the ACMG/AMP variant classification guidelines and standard bioinformatic data formats (e.g. BAM, VCF, etc.)—the training grounded technical scripts in actionable clinical utility. While traditional bioinformatics training often demands steep programming prerequisites, our results introduce a novel paradigm: AI can act as a robust supporting tool that lowers technical barriers, allowing non-technical healthcare professionals to achieve applied technical competencies within compressed timeframes.

From a health systems perspective, these findings offer highly scalable workforce development strategies for precision medicine, directly supporting Saudi Arabia's Vision 2030 mandate and its

Health Sector Transformation Program. As the Kingdom rapidly scales its genomic medicine infrastructure, a primary bottleneck is workforce readiness. To build upon this baseline, future educational interventions should transition from a one-size-fits-all model to customized, modular tracks: patient-facing professionals require more clinical case studies, whereas non-patient-facing professionals need advanced technical programming. Future iterations must expand ethical, legal, and social implications (ELSI) training and incorporate longitudinal online modules to measure long-term application. Finally, forming strategic partnerships with technology companies developing healthcare-specific AI systems will ensure curriculum content remains cutting-edge and deeply integrated with local healthcare contexts.

Limitations

Several technical and design limitations must be acknowledged when evaluating these outcomes. First, due to a technical glitch in the pre-survey deployment, the pre- and post-workshop surveys could not be linked or paired at the individual participant level. Consequently, the datasets had to be treated as independent samples rather than paired observations, which precluded a direct intra-individual tracking of confidence trajectories. To mitigate this design flaw, we applied a conservative statistical analysis using non-paired comparative testing; this approach minimizes the risk of Type I errors (false positives), ensuring that the observed confidence gains are statistically robust despite the lack of individual pairing.

Second, the overall sample size—particularly regarding the post-workshop response rate—was modest, which inherently limits the statistical power of the comparative analyses and restricts the broad generalizability of the findings. Third, the study relied strictly on self-reported confidence and attitudes rather than objective, external metrics of bioinformatics skill acquisition (such as practical coding tests) or direct clinical performance tracking. This reliance introduces a potential social desirability bias, where participants might overestimate their perceived proficiency to align with instructor expectations.

Furthermore, because the workshop was promotional and optional, the cohort was entirely self-selected without an external control group for baseline comparison. It is highly probable that these self-selected attendees possessed a greater baseline interest and enthusiasm for Clinical Bioinformatics and AI than the broader, general healthcare workforce. Finally, the compressed 3-day duration of the course structurally limited the absolute depth of programming and clinical skill acquisition, forcing instructors to dynamically reduce manual programming sessions in favor of generative AI script production. This constraint also restricted critical systemic discussions, such as the ELSI of genomic medicine, to a single introductory presentation. While these combined constraints mean the results must be interpreted as a localized proof-of-concept, the real-time feedback loops via site surveys and mentor interactions provided a robust framework for immediate course refinement.

Conclusions

Despite its inherent design limitations, our workshop and evaluation study successfully highlighted a rapidly growing interest among healthcare professionals in Clinical Bioinformatics and AI. The study clearly demonstrated a significant increase in participants' self-confidence regarding both their foundational Clinical Bioinformatics knowledge and their practical ability to develop data analytic applications and tools. Additionally, the structured, balanced introduction of AI was highly effective in clarifying immediate clinical use cases while simultaneously underscoring the critical importance of caution, rigorous testing, and responsible implementation in active clinical practice. The observed effect sizes and consistent directional trends strongly support the educational value of this short-form intervention. Ultimately, this exploratory study provides a solid rationale and framework for larger, multi-institutional longitudinal studies that incorporate objective performance metrics to track long-term clinical impacts and patient outcomes.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with national and international regulations governing research involving human participants, including the Declaration of Helsinki and national Saudi regulations. This study was approved and cleared by the Institutional Review Board (IRB) of King Faisal Specialist Hospital and Research Centre (Clearance number: 2265467). The requirement for written informed consent from participants was waived; however, the online survey included an informed consent statement describing that participation was voluntary and anonymous.

Consent for publication

Not applicable

Availability of data and materials

Survey responses and summary statistics are available in the supplementary files.

Competing interests

MAL and AAB declare that they were the scientific co-chairs of the training workshop being evaluated in this study, which represents a potential non-financial competing interest. The authors maintained objectivity in the data collection and analysis.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The workshop was conducted using internal institutional resources.

Authors' contributions

Conceptualization: MAL (lead), AAB (equal), MM (equal), CBAIH 2025 Workshop Team (supporting); *Data curation:* AF (lead), MAL (equal), MAB (equal); *Formal analysis:* MAL (lead), MAB (supporting); *Funding acquisition:* AAL (lead), SB (supporting); *Investigation:* AF (lead), MAL (equal), MAB (supporting), MM (equal); *Methodology:* MAL (lead), MAB (supporting), MM (equal), AF (supporting); *Project*

administration: MAL (lead), AF (equal), AAB (supporting); *Resources:* AAL (lead), SB (supporting); *Supervision:* MAL (lead), AAB (equal), AAL (equal), SB (supporting); *Validation:* MAL (lead), AAB (equal); *Visualization:* MAL (lead), MAB (supporting); *Writing – original draft:* MAB (lead), AF (equal), MM (supporting), MAL (supporting); *Writing – review & editing:* MAL (lead), AF (equal), MM (equal), MAB (equal), AAB (equal), AAL (equal), SB (equal).

Acknowledgement

The authors would like to thank all participants of the inaugural Clinical Bioinformatics and AI in Healthcare (CBAIH) workshop for their valuable contributions to the study survey. We also acknowledge the support of the Academic & Training Affairs (ATA) at King Faisal Specialist Hospital and Research Centre (KFSH&RC) for facilitating the delivery of the workshop. In addition, we extend our sincere appreciation to the CBAIH2025 Workshop Team (organizing committee members and speakers) for their contributions to the development and implementation of the program, including Anas Alazami (KFSH&RC), Wassim Hossari (KFSH&RC), Ruqaiah Altassan(KFSH&RC), Nivine Alakili (KFSH&RC), Wedad Albalawi (KFSH&RC), Meshari Alwazae (KFSH&RC), Tareg Alghamdi (KFSH&RC), Walaa Alenazi (KFSH&RC), Hala Almutiri (KFSH&RC), Sarah Alhamad (KFSH&RC), Sarah Abdulrahman (KFSH&RC), Mohammed Alsayed (KFSH&RC), Petros Kotsidis (KFSH&RC), Mohammed Bamajboor (KFSH&RC), Hanin Omer (KFSH&RC), Wail Baalawi (LiferaOmics), Abdullah Alfalah (KFSH&RC), Maddirevula Sateesh (KFSH&RC), Rola Qarrash (KFSH&RC), Turki Alzahrani (KFSH&RC), Abdullah Alsuwaidan (KFSH&RC), Mohammed Alowain (KFSH&RC), Azza Althagafi (Taif University), Mohamed Elkalioby (KFSH&RC), Asma Tahir (KFSH&RC), and Zeeshan Shah (KFSH&RC).

References

- [1] J. D. Tenenbaum, "Translational Bioinformatics: Past, Present, and Future," *Genomics, Proteomics & Bioinformatics*, vol. 14, no. 1, pp. 31–41, Feb. 2016, doi: 10.1016/j.gpb.2016.01.003.
- [2] S. Singh, A. K. Pandey, and V. K. Prajapati, "From genome to clinic: The power of translational bioinformatics in improving human health," *Adv Protein Chem Struct Biol*, vol. 139, pp. 1–25, 2024, doi: 10.1016/bs.apcsb.2023.11.010.
- [3] R. Bellazzi, M. Masseroli, S. Murphy, A. Shabo, and P. Romano, "Clinical Bioinformatics: challenges and opportunities," *BMC Bioinformatics*, vol. 13 Suppl 14, no. Suppl 14, p. S1, 2012, doi: 10.1186/1471-2105-13-S14-S1.
- [4] S. Kadri *et al.*, "Clinical Bioinformatician Body of Knowledge—Bioinformatics and Software Core: A Report of the Association for Molecular Pathology," *The Journal of Molecular Diagnostics*, vol. 27, no. 7, pp. 566–582, Jul. 2025, doi: 10.1016/j.jmoldx.2025.04.008.
- [5] M. Abouelhoda, "The emerging field of clinical Bioinformatics: Infrastructure, analytical tools, and knowledge bases," in *2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME)*, Oct. 2017, pp. 1–4. doi: 10.1109/ICABME.2017.8167566.
- [6] F. Carrasco-Ramiro, R. Peiró-Pastor, and B. Aguado, "Human genomics projects and precision medicine," *Gene Ther*, vol. 24, no. 9, pp. 551–561, Sep. 2017, doi: 10.1038/gt.2017.77.
- [7] J. S. Nunn, J. Tiller, P. Fransquet, and P. Lacaze, "Public Involvement in Global Genomics Research: A Scoping Review," *Front. Public Health*, vol. 7, Apr. 2019, doi: 10.3389/fpubh.2019.00079.
- [8] All of Us Research Program Investigators *et al.*, "The 'All of Us' Research Program," *N Engl J Med*, vol. 381, no. 7, pp. 668–676, Aug. 2019, doi: 10.1056/NEJMSr1809937.
- [9] H. Ateia *et al.*, "Population Genome Programs across the Middle East and North Africa: Successes, Challenges, and Future Directions," *Biomed Hub*, vol. 8, no. 1, pp. 60–71, Sep. 2023, doi: 10.1159/000530619.
- [10] I. Slade, D. N. Subramanian, and H. Burton, "Genomics education for medical professionals – the current UK landscape," *Clinical Medicine*, vol. 16, no. 4, pp. 347–352, Aug. 2016, doi: 10.7861/clinmedicine.16-4-347.
- [11] Z. Stark *et al.*, "Integrating Genomics into Healthcare: A Global Responsibility," *Am J Hum Genet*, vol. 104, no. 1, pp. 13–20, Jan. 2019, doi: 10.1016/j.ajhg.2018.11.014.
- [12] B. Yan, K. H. K. Ban, and T. W. Tan, "Integrating translational bioinformatics into the medical curriculum," *Int J Med Educ*, vol. 5, pp. 132–134, Jun. 2014, doi: 10.5116/ijme.53ae.bc97.
- [13] D. Talwar, T.-S. Tseng, M. Foster, L. Xu, and L.-S. Chen, "Genetics/genomics education for nongenetic health professionals: a systematic literature review," *Genet Med*, vol. 19, no. 7, pp. 725–732, Jul. 2017, doi: 10.1038/gim.2016.156.
- [14] A. Nisselle, B. Terrill, M. Janinski, S. Metcalfe, and C. Gaff, "Ensuring best practice in genomics education: A scoping review of genomics education needs assessments and evaluations," *Am J Hum Genet*, vol. 111, no. 8, pp. 1508–1523, Aug. 2024, doi: 10.1016/j.ajhg.2024.06.005.
- [15] C. Wray *et al.*, "The HUGO Clinical Genomics & Genomic Medicine Education Survey: clinicians globally need and want genomic medicine training," *Hum Genomics*, vol. 19, p. 139, Nov. 2025, doi: 10.1186/s40246-025-00841-7.
- [16] G. López-Campos, V. López-Alonso, and F. Martín-Sánchez, "Training Health Professionals in Bioinformatics," *Methods Inf Med*, vol. 49, no. 3, pp. 299–304, 2010, doi: 10.3414/ME09-02-0008.
- [17] M. D. Brazas, F. Lewitter, M. V. Schneider, C. W. G. van Gelder, and P. M. Palagi, "A Quick Guide to Genomics and Bioinformatics Training for Clinical and Public Audiences," *PLoS Comput Biol*, vol. 10, no. 4, p. e1003510, Apr. 2014, doi: 10.1371/journal.pcbi.1003510.
- [18] B. Yan, K. H. K. Ban, and T. W. Tan, "Integrating translational bioinformatics into the medical curriculum," *Int J Med Educ*, vol. 5, pp. 132–134, Jun. 2014, doi: 10.5116/ijme.53ae.bc97.

- [19] "Developing Clinical Bioinformatics Training in the NHS," Genomics Education Programme. Accessed: Jan. 15, 2026. [Online]. Available: <http://www.genomicseducation.hee.nhs.uk/documents/developing-clinical-bioinformatics-training-in-the-nhs/>
- [20] N. Mulder *et al.*, "The development and application of bioinformatics core competencies to improve bioinformatics training and education," *PLOS Computational Biology*, vol. 14, no. 2, p. e1005772, Feb. 2018, doi: 10.1371/journal.pcbi.1005772.
- [21] H. Naveed *et al.*, "A Comprehensive Overview of Large Language Models," *ACM Trans. Intell. Syst. Technol.*, vol. 16, no. 5, p. 106:1-106:72, Aug. 2025, doi: 10.1145/3744746.
- [22] "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *SciSpace - Paper*, Association for Computational Linguistics, Oct. 2018, pp. 4171–4186. doi: 10.18653/V1/N19-1423.
- [23] D. Idan and S. Einav, "Primer on large language models: an educational overview for intensivists," *Crit Care*, vol. 29, p. 238, Jun. 2025, doi: 10.1186/s13054-025-05479-4.
- [24] S. Johnson and D. Hyland-Wood, "A Primer on Large Language Models and their Limitations," *Qeios*, Jan. 2025, doi: 10.32388/NHJYVS.
- [25] H. Touvron *et al.*, "LLaMA: Open and Efficient Foundation Language Models," Feb. 27, 2023, *arXiv*: arXiv:2302.13971. doi: 10.48550/arXiv.2302.13971.
- [26] S. M. S. Mohammadabadi *et al.*, "A Survey of Large Language Models: Evolution, Architectures, Adaptation, Benchmarking, Applications, Challenges, and Societal Implications," *Electronics*, vol. 14, no. 18, Sep. 2025, doi: 10.3390/electronics14183580.
- [27] "Multi-Modal Generative AI: Multi-modal LLM, Diffusion and Beyond." Accessed: Jan. 19, 2026. [Online]. Available: <https://arxiv.org/html/2409.14993v1>
- [28] C. Chen *et al.*, "Diffusion Models For Multi-Modal Generative Modeling," Sep. 24, 2024, *arXiv*: arXiv:2407.17571. doi: 10.48550/arXiv.2407.17571.
- [29] "ShareGPT-4o-Image: Aligning Multimodal Models with GPT-4o-Level Image Generation." Accessed: Jan. 19, 2026. [Online]. Available: <https://arxiv.org/html/2506.18095v1>
- [30] Z. Xi *et al.*, "The Rise and Potential of Large Language Model Based Agents: A Survey," Sep. 19, 2023, *arXiv*: arXiv:2309.07864. doi: 10.48550/arXiv.2309.07864.
- [31] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative Agents: Interactive Simulacra of Human Behavior," Aug. 06, 2023, *arXiv*: arXiv:2304.03442. doi: 10.48550/arXiv.2304.03442.
- [32] M. Abou Ali, F. Dornaika, and J. Charafeddine, "Agentic AI: a comprehensive survey of architectures, applications, and future directions," *Artif Intell Rev*, vol. 59, no. 1, p. 11, Nov. 2025, doi: 10.1007/s10462-025-11422-4.
- [33] K. Kang and M. A. P. Bagaoisan, "Research Status of the Knowledge-Attitude-Practice Theory Model in Gastric Cancer Prevention," *Cureus*, vol. 16, no. 7, p. e64960, Jul. 2024, doi: 10.7759/cureus.64960.
- [34] Alfares, A., Imtiaz, F., Maddirevula, S. *et al.* "Building genomic medicine in Saudi Arabia". *Nature Genetics*, 58, 463–466 (2026). doi:10.1038/s41588-026-02513-1
- [35] Anderson, Lorin W., and David R. Krathwohl, eds. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison Wesley Longman, Inc.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary1CBAIHWorshopSurveyBlinded.xlsx](#)
- [Supplementary2InfoSurveyQuestionsStats.xlsx](#)