

ChatGPT και Εφαρμογές AI για Ιατρούς

4th session – Κατάλληλα ChatGPT Prompts για Ιατρούς, Clustering and Genetic Algorithms

UNIVERSITY OF THE
AEGEAN



SCHOOL OF ENGINEERING
DEPARTMENT OF INFORMATION
AND COMMUNICATION
SYSTEMS ENGINEERING

Presenter: Panagiotis Symeonidis

Associate Professor

<http://panagiotissymeonidis.com>

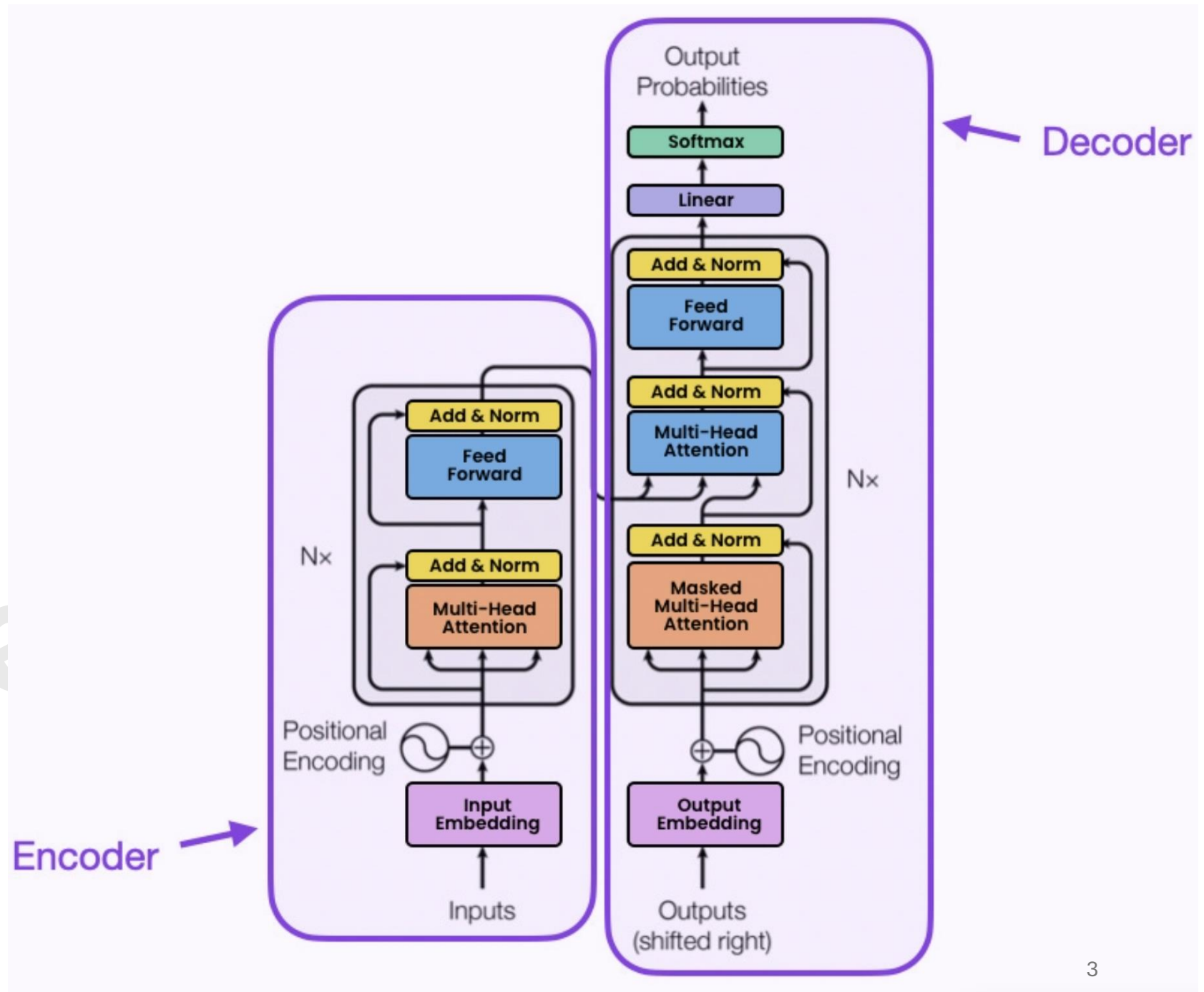
psymeon@aegean.gr

what does chatGPT stand for?

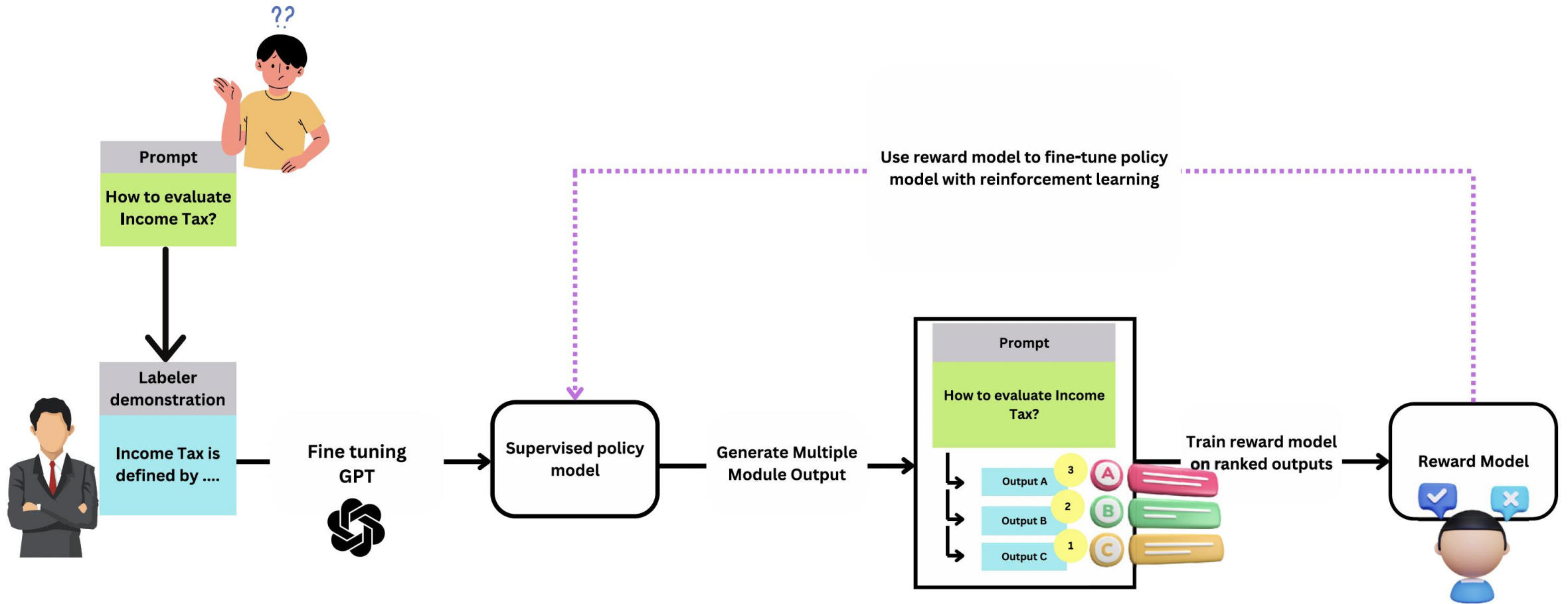
- GPT stands for **Generative Pre-trained Transformer**

- **Generative** because it can generate content
- **Pre-trained** because it is based on a language prediction model, which is trained off-line.
- **Transformer** because it performs Word2vec and Vec2Word operations

Transformer Architecture

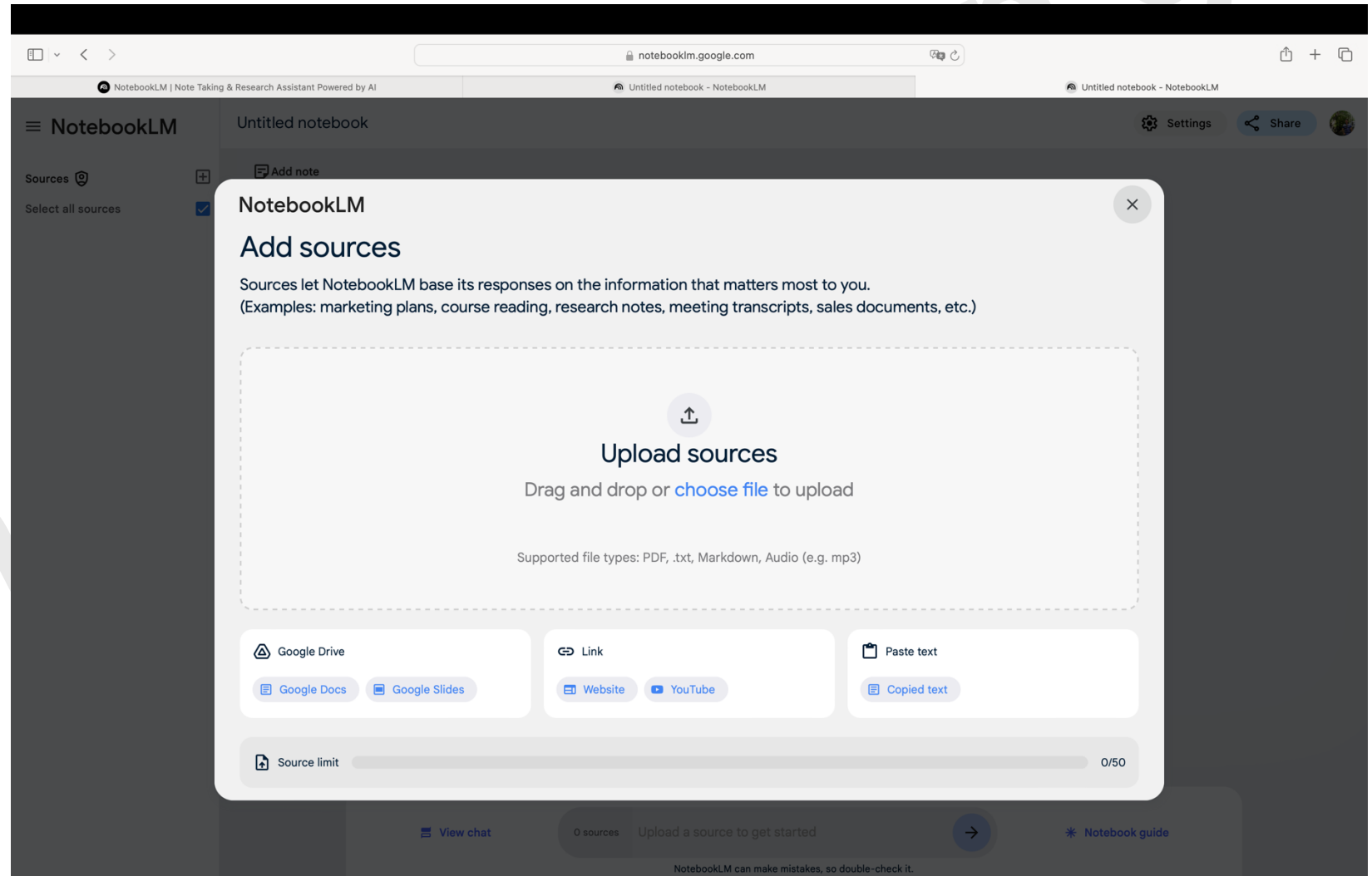


ChatGPT training process



Google's NotebookLM uses RAG technology

You upload pdfs and it provides answers to your questions from the content of the documents you uploaded



General ChatGPT Prompts

Θέλω να πάρεις τον ρόλο ενός εκπαιδευτή ιατρών που θα εκπαιδεύσει τους ιατρούς στην χρήση του ChatGPT 4o. Φτιάξε 10 διαφάνειες για βέλτιστες πρακτικές για κατάλληλα ChatGPT prompts που πρέπει να γνωρίζουν οι ιατροί και τι μπορούν να ζητούν από το ChatGPT 4o ώστε να βελτιώνουν την παραγωγικότητα τους. Μία από τις 10 διαφάνειες θα μπορούσε να αφορά τον καθορισμό περιορισμών στα ChatGPT prompts. Μια άλλη θα μπορούσε να αφορά τα παιχνίδια ρόλων, κτλ.



Διαφάνεια 1: Εισαγωγή στο ChatGPT-4 και τη χρήση του στην Ιατρική

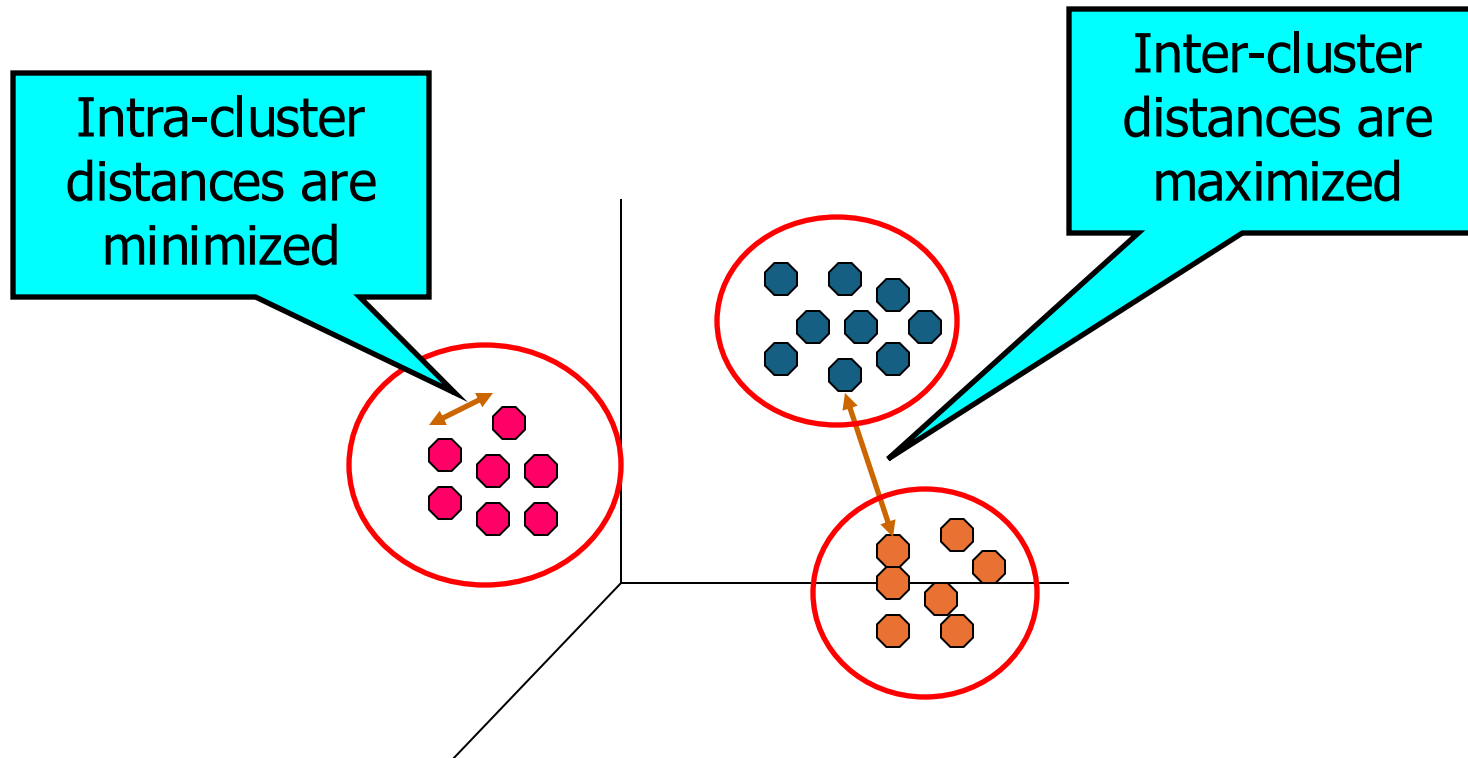
- **Τι είναι το ChatGPT-4:** Ένα εξελιγμένο μοντέλο γλωσσικής επεξεργασίας που χρησιμοποιείται για διάλογο, έρευνα και ανάλυση πληροφοριών.
- **Πλεονεκτήματα στη χρήση στην ιατρική:** Ταχεία αναζήτηση πληροφοριών, υποστήριξη στη διάγνωση, εκπαίδευση και βελτίωση της παραγωγικότητας.
- **Στόχος της εκπαίδευσης:** Να μάθετε πώς να αξιοποιείτε το ChatGPT για να εξοικονομήτε χρόνο και να βελτιώνετε τις κλινικές αποφάσεις.

Διαφάνεια 2: Βέλτιστες Πρακτικές για Σαφή Prompts

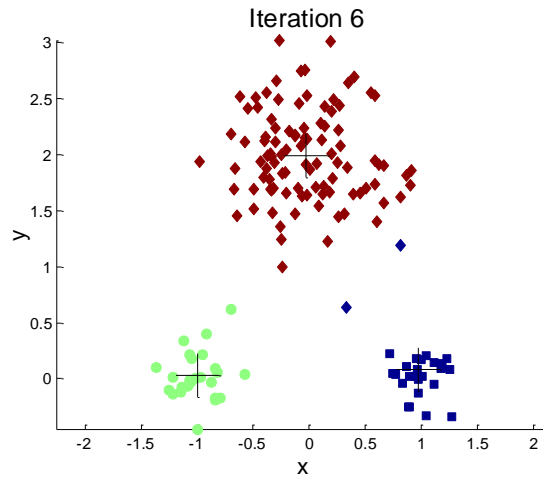
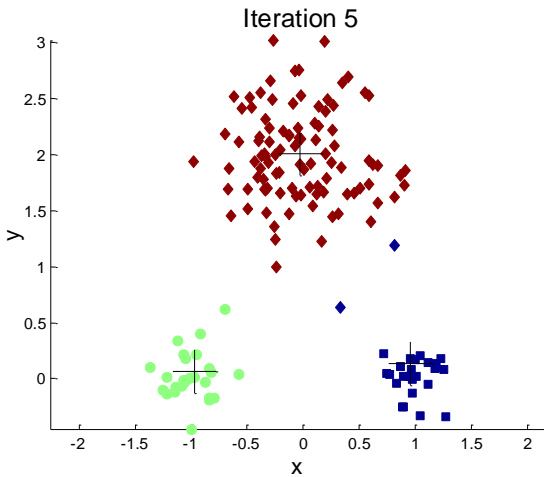
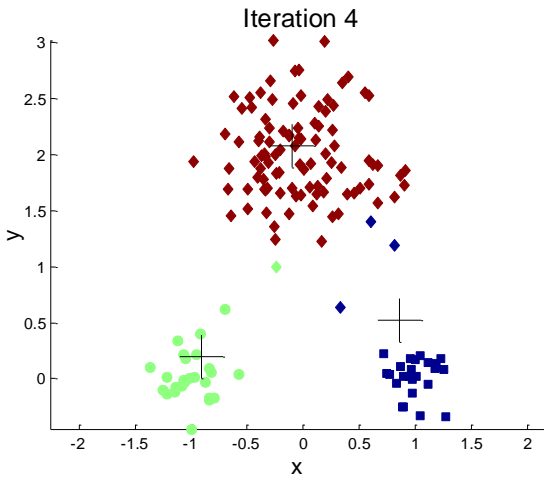
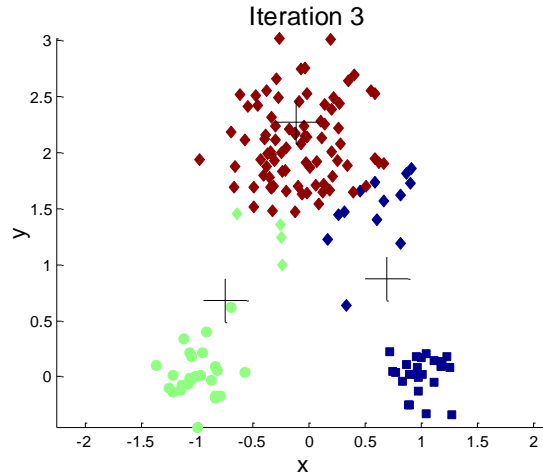
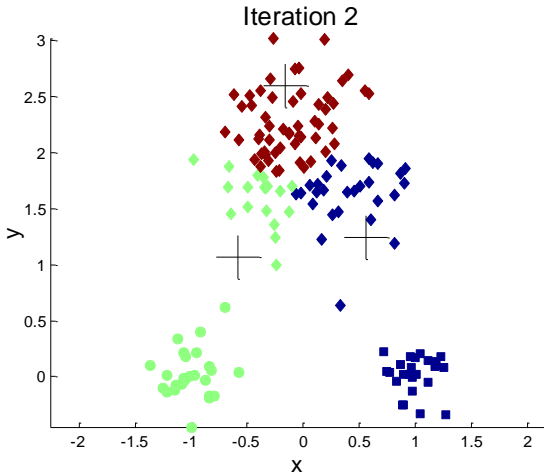
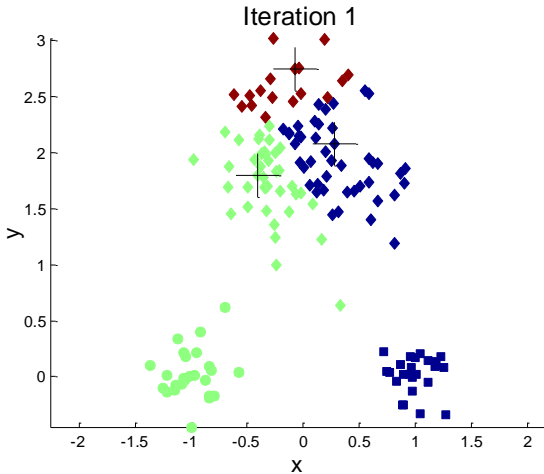
- **Δώστε σαφείς οδηγίες:** Προσδιορίστε το ερώτημα με σαφήνεια και περιγράψτε τι ακριβώς χρειάζεστε από το ChatGPT.
- **Χρήση σύντομων και ακριβών περιγραφών:** Αποφύγετε την περιττή πολυπλοκότητα στις ερωτήσεις σας.
- **Παραδείγματα:**
 - "Ποιες είναι οι πιο πρόσφατες κατ[...]
↓
νητήριες οδηγίες για τη θεραπεία της υπέρτασης;"

What is Cluster Analysis?

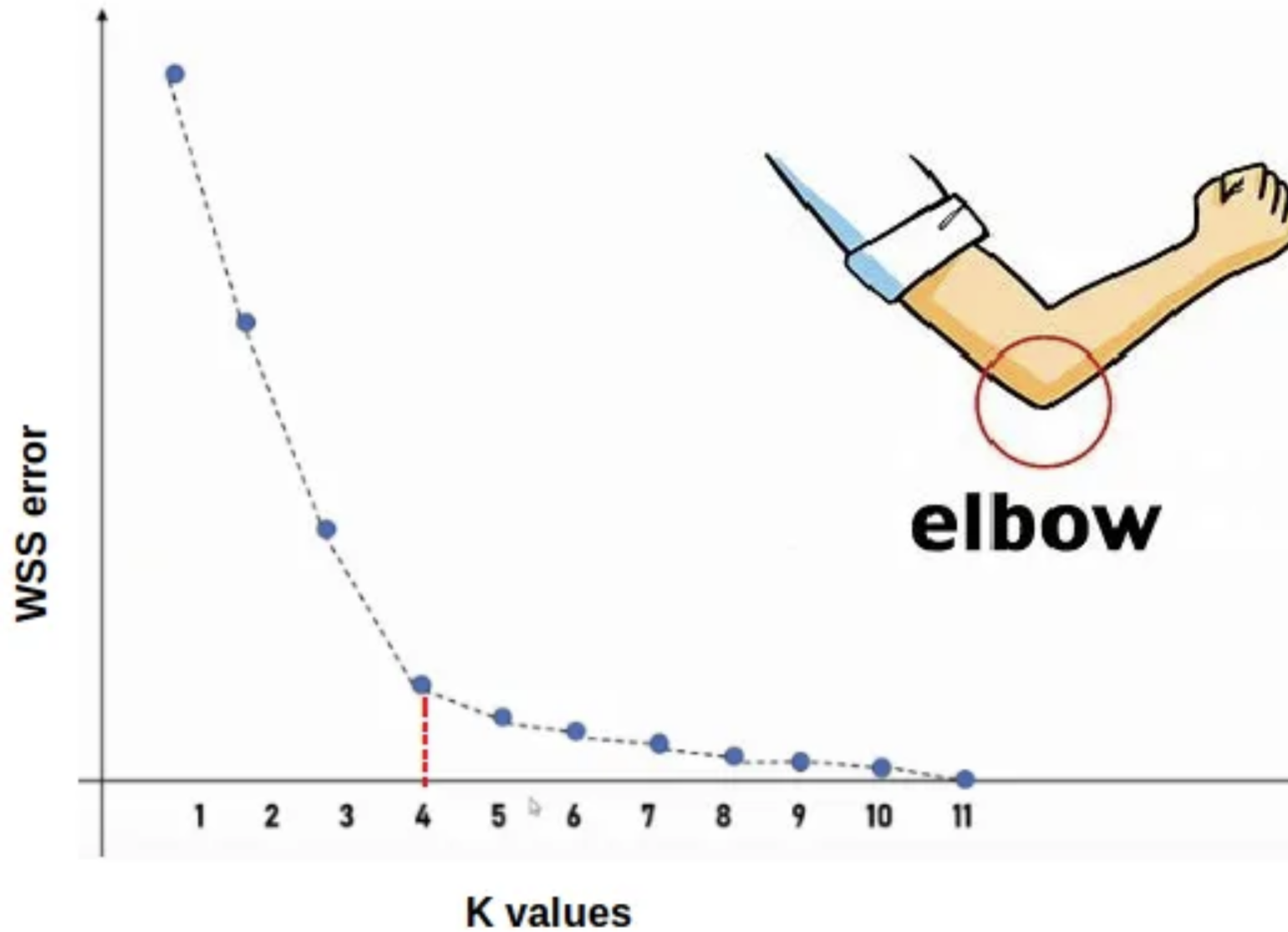
- Given a set of objects, place them in groups such that the objects in a group are similar (or related) to one another and different from (or unrelated to) the objects in other groups



Example of K-means Clustering



Elbow method

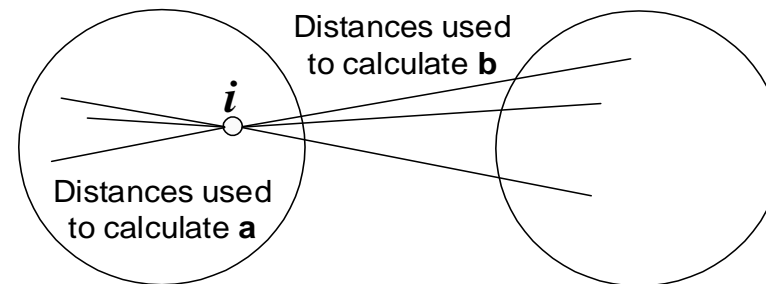


Silhouette Coefficient

- Silhouette coefficient combines ideas of both cohesion and separation, but for individual points, as well as clusters and clusterings
- For an individual point, i
 - Calculate a = average distance of i to the points in its cluster
 - Calculate b = min (average distances of i to points in all other cluster)
 - The silhouette coefficient for a point is then given by

$$s = (b - a) / \max(a, b)$$

- Value can vary between -1 and 1
- Typically ranges between 0 and 1.
- The closer to 1 the better.



For n_clusters = 2 The average silhouette_score is :
0.9419743880621418

For n_clusters = 3 The average silhouette_score is :
0.8925568467675032

For n_clusters = 4 The average silhouette_score is :
0.8854468255579183

For n_clusters = 5 The average silhouette_score is :
0.8859344049988384

For n_clusters = 6 The average silhouette_score is :
0.896222949688388

For n_clusters = 7 The average silhouette_score is :
0.9531228433846561

For n_clusters = 8 The average silhouette_score is :
0.9882303235394505

For n_clusters = 9 The average silhouette_score is :
0.9942722572401562

For n_clusters = 10 The average silhouette_score is :
0.9860105575225317

**For n_clusters = 11 The average silhouette_score is :
0.9999999922837097**

For n_clusters = 12 The average silhouette_score is :
0.906560275971653

• **Cluster 0: Current Smokers with Moderate Health Risks**

- **Age:** Approximately 44 years
- **Hypertension:** Low prevalence (7.95%)
- **Heart Disease:** Very low prevalence (3.34%)
- **BMI:** Around 28.4
- **HbA1c Level:** Around 5.54
- **Blood Glucose Level:** Around 139.5
- **Diabetes:** Low prevalence (9.85%)
- **Smoking History:** All individuals are current smokers

• **Cluster 1: Non-smokers with Low Health Risks**

- **Age:** Approximately 42 years
- **Hypertension:** Very low prevalence (0%)
- **Heart Disease:** Very low prevalence (0%)
- **BMI:** Around 27.8
- **HbA1c Level:** Around 5.51
- **Blood Glucose Level:** Around 137.2
- **Diabetes:** Low prevalence (7.23%)
- **Smoking History:** Majority never smoked (77.41%), with some having a history of smoking

• **Cluster 2: Younger Individuals with Minimal Health Issues**

- **Age:** Approximately 32 years
- **Hypertension:** Very low prevalence (0%)
- **Heart Disease:** Very low prevalence (0%)
- **BMI:** Around 25.1
- **HbA1c Level:** Around 5.45
- **Blood Glucose Level:** Around 134.6
- **Diabetes:** Very low prevalence (3.04%)
- **Smoking History:** Majority have no information on smoking history (100%)

• **Cluster 3: Former Smokers with Moderate to High Health Risks**

- **Age:** Approximately 57 years
- **Hypertension:** Higher prevalence (12.25%)
- **Heart Disease:** Moderate prevalence (7.53%)
- **BMI:** Around 29.6
- **HbA1c Level:** Around 5.64
- **Blood Glucose Level:** Around 142.9
- **Diabetes:** Higher prevalence (16.32%)
- **Smoking History:** All individuals are former smokers

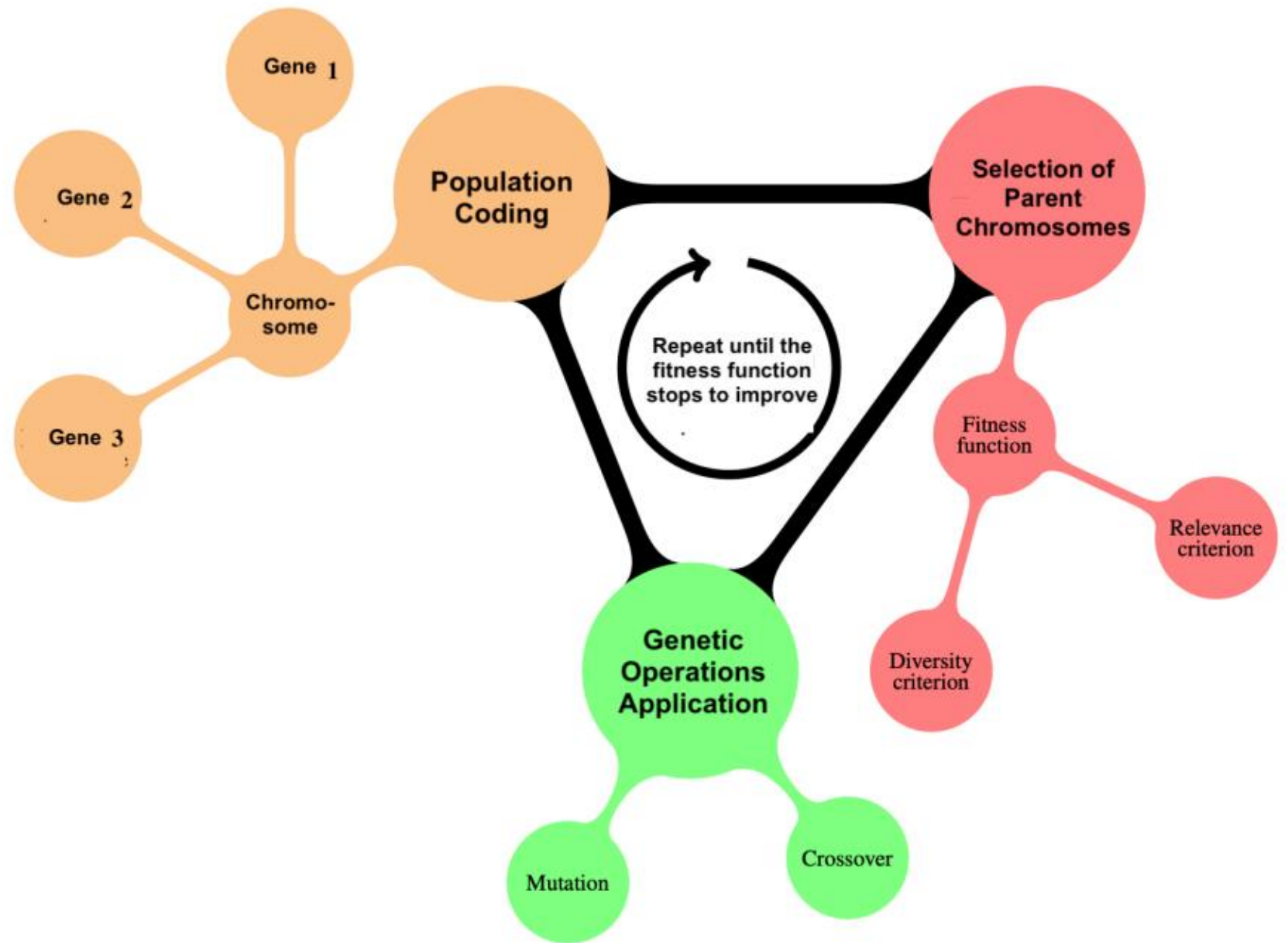
• **Cluster 4: Older Individuals with High Health Risks**

- **Age:** Approximately 64 years
- **Hypertension:** Very high prevalence (73.51%)
- **Heart Disease:** High prevalence (38.43%)
- **BMI:** Around 30.3
- **HbA1c Level:** Around 5.83
- **Blood Glucose Level:** Around 150.2
- **Diabetes:** Higher prevalence (28.29%)
- **Smoking History:** Mixed smoking history with a significant proportion never smoked (52.09%) and some having a history of smoking

Genetic Algorithms

- The genetic algorithms exploit important mechanisms of the natural functions of organisms
 - selection,
 - crossover,
 - mutation
- Genetic algorithms follow a search process for the optimal solution, which is guided by a **fitness function**, that evaluates a large number of different possible solutions.

Architecture of a genetic algorithm



Algorithm 4 Optimizing the neighborhood of the nearest users.

Input:

$U = \{u_1, u_2, \dots, u_n\}$: the set of users.

u_0 : the target user.

m : a predefined maximum number of generations to be generated.

f : the size of the initial population.

Output:

N_0 : the optimal neighborhood of user u_0 or else the optimal chromosome.

1: population \leftarrow Creating a random population (size = f)

2: **While** the number of generations $g < m$ **do**

 descendants \leftarrow applySelectionOperator(population)

Parent Selection

 descendants \leftarrow applyCrossoverOperator(descendants)

Crossover

 descendants \leftarrow applyMutationOperator(descendants)

Mutation

 Population \leftarrow descendants \cup Population

$g \leftarrow g+1$

3: **end_while**

4: $N_0 \leftarrow$ Select_best_neighborhood (population) {based on their Fitness score}

Genetic Algorithms