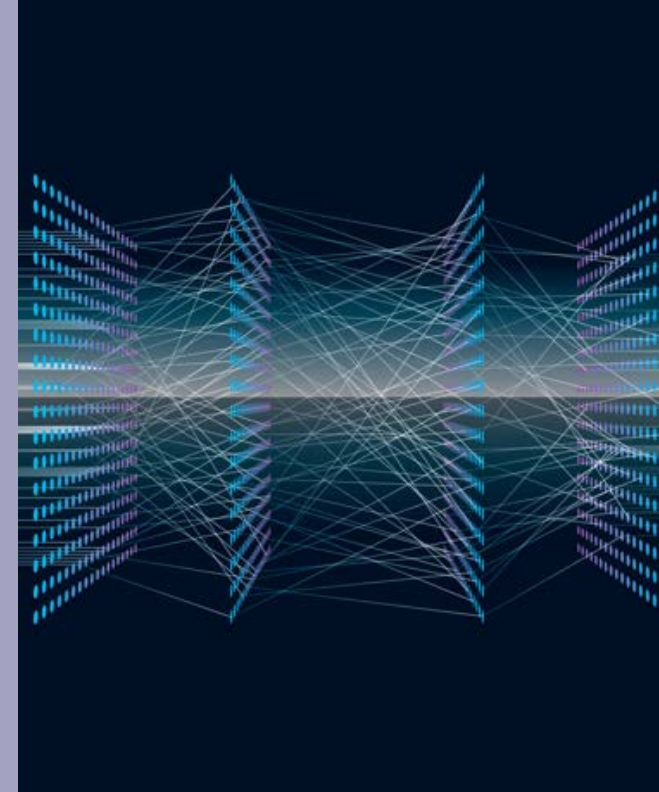




Presentation Title

Δημήτρης Κυριάκου





Εισαγωγή

Σκοπός της διπλωματικής εργασίας ήταν η μελέτη του συνδυασμού των τεχνολογιών της **Μηχανικής Μάθησης** και του **Blockchain**. Στα πλαίσια της εργασίας δημιουργήθηκε μία εφαρμογή που αξιοποιεί συστήματα Blockchain για την εκπαίδευση και βελτίωση μοντέλων μηχανικής μάθησης

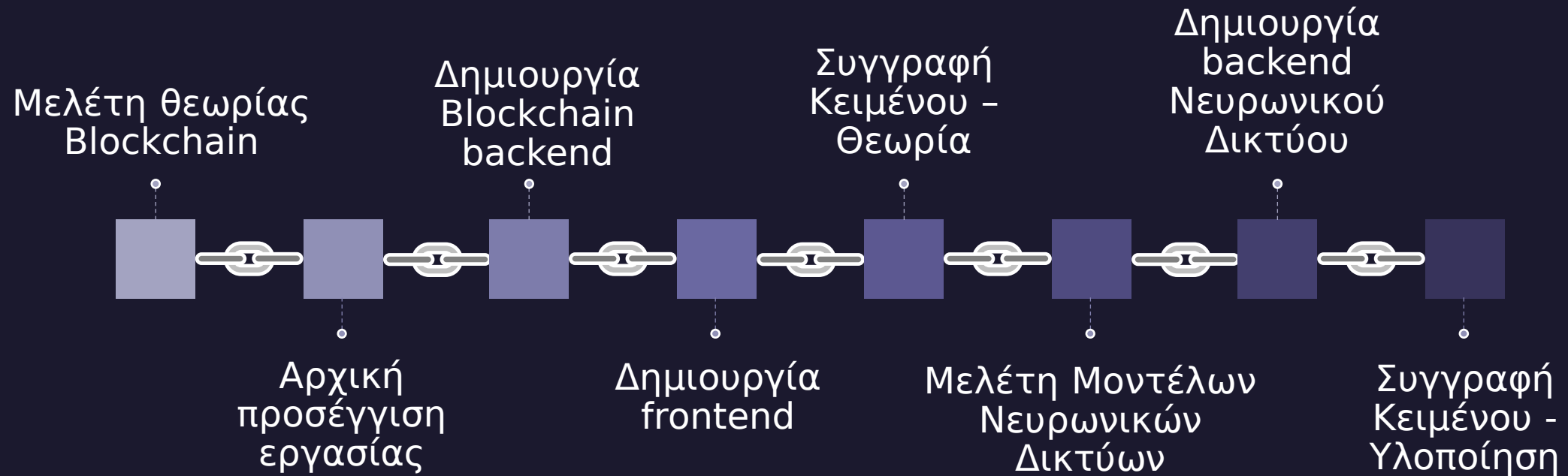
Περιεχόμενα

1. Χρονοδιάγραμμα εργασίας
2. Βασικές Έννοιες
3. Εφαρμογή DEMOS
4. Παρουσίαση εφαρμογής
5. Συμπεράσματα



1. Χρονοδιάγραμμα εργασίας

Χρονοδιάγραμμα εργασίας



2. Βασικές Έννοιες

Τι είναι Blockchain

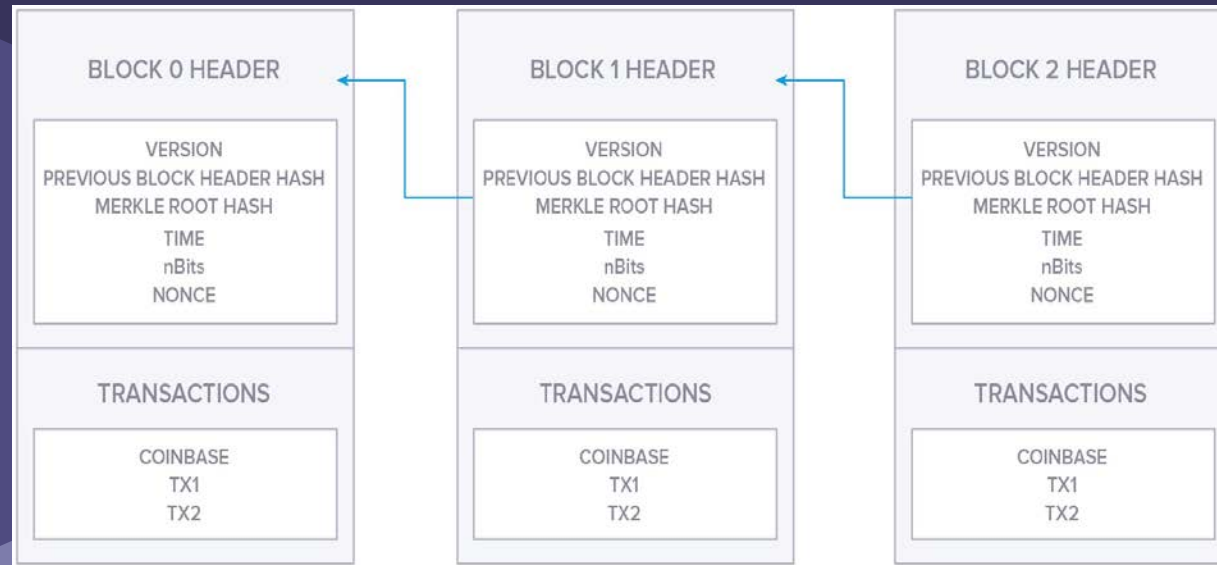
Τι είναι Ethereum

Blockchain και Μηχανική Μάθηση

Blockchain

Το Blockchain είναι ένα **αμετάβλητο κατανεμημένο καθολικό**
(immutable distributed ledger)

Αποτελείται από blocks
τα οποία περιέχουν
δεδομένα (στοιχεία
συναλλαγών) και είναι
ενωμένα
δημιουργώντας μια
αλυσίδα



Το δίκτυο Blockchain
αποτελείται από
υπολογιστικές
μονάδες που
εκτελούν,
καταγράφουν και
επιβεβαιώνουν
συναλλαγές

Υπάρχει ένα πρωτόκολλο συναίνεσης (consensus protocol) που εξασφαλίζει ότι η κατάσταση του δικτύου είναι μοναδική και κοινώς αποδεκτή



Ethereum

Το Ethereum είναι ένα Blockchain που διαφέρει από τα υπόλοιπα



Προσφέρει δυνατότητα προγραμματισμού χρησιμοποιώντας μια Turing Complete γλώσσα για τη δημιουργία Smart Contracts



Συνδυάζοντας Smart Contracts με περιβάλλον διεπαφής μπορούν να δημιουργηθούν αποκεντρωμένες εφαρμογές (dApps - Decentralized Applications)

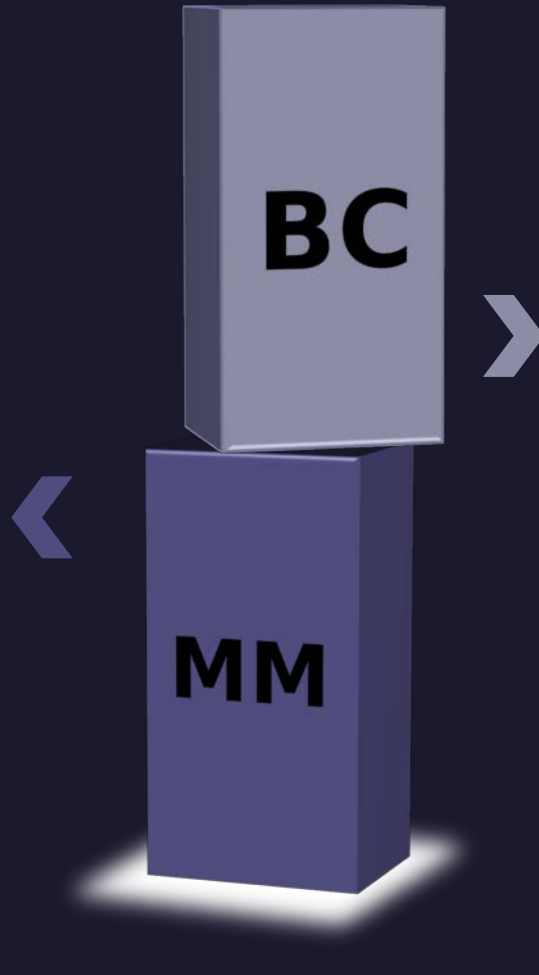


Τα Smart Contracts είναι προγράμματα στο δίκτυο του Ethereum που μπορούν να εκτελεστούν από τους κόμβους

Μηχανική Μάθηση και Blockchain

Μηχανική Μάθηση

- Πολλές και απαιτητικές πράξεις
- Απαίτηση για μεγάλο όγκο δεδομένων
- Απαίτηση ποιοτικών δεδομένων



Blockchain

- Καταμερισμός εργασίας
- Παραγωγή μεγάλου όγκου δεδομένων
- Συνεχής έλεγχος ποιότητας δεδομένων

3. Εφαρμογή DEMOS

Πρόβλημα

Κεντρική ιδέα

Τύποι χρηστών

Μηχανισμός Κινήτρου

Μοντέλο Νευρωνικού Δικτύου

Υλοποίηση

Πρόβλημα

Στόχος κάθε μοντέλου Μηχανικής Μάθησης είναι η βελτίωση της απόδοσής του. Μέσα από αυτή τη διπλωματική εργασία διερευνήθηκαν τρόποι με τους οποίους θα μπορούσαν να αξιοποιηθούν συστήματα Blockchain για την εκπαίδευση και βελτίωση τέτοιων μοντέλων Μηχανικής Μάθησης.

Μία κύρια μέθοδος βελτίωσης της απόδοσης είναι μέσω του εμπλουτισμού του συνόλου δεδομένων εκπαίδευσης. Η συλλογή δεδομένων εκπαίδευσης είναι μια απλή αλλά χρονοβόρα διαδικασία και είναι απαραίτητο να γίνει καλή αξιολόγηση των δεδομένων. Αυτό την καθιστά δύσκολη όταν γίνεται από ένα και μόνο χρήστη.

Ως εκ τούτου δημιουργήθηκε το **DEMOS** (Distributedly Enhanced Machine learning Optimization System) ή αλλιώς **Κατανεμημένα Ενισχυμένο Σύστημα Βελτιστοποίησης Μηχανικής Μάθησης**

Το DEMOS χρησιμοποιείται για τη συνεχή συλλογή δεδομένων εκπαίδευσης, την επανεκπαίδευση μοντέλων ΝΔ αλλά και την επαναξιολόγησή τους

Κεντρική ιδέα

DEMOS



1

Ένας κόμβος αιτείται τη συλλογή δεδομένων εκπαίδευσης για κάποιο μοντέλο ΝΔ επιβλεπόμενης μάθησης



2

Οι κόμβοι του δικτύου προσφέρουν και επαληθεύουν νέα δεδομένα εκπαίδευσης

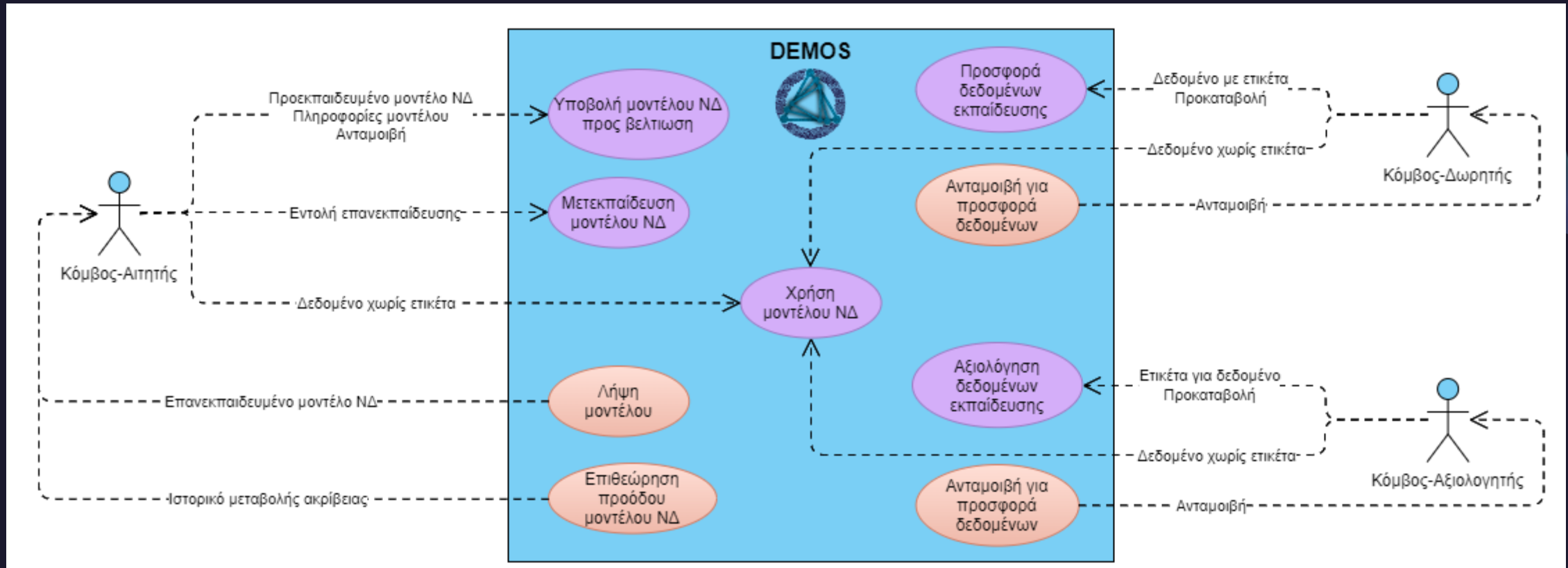


3

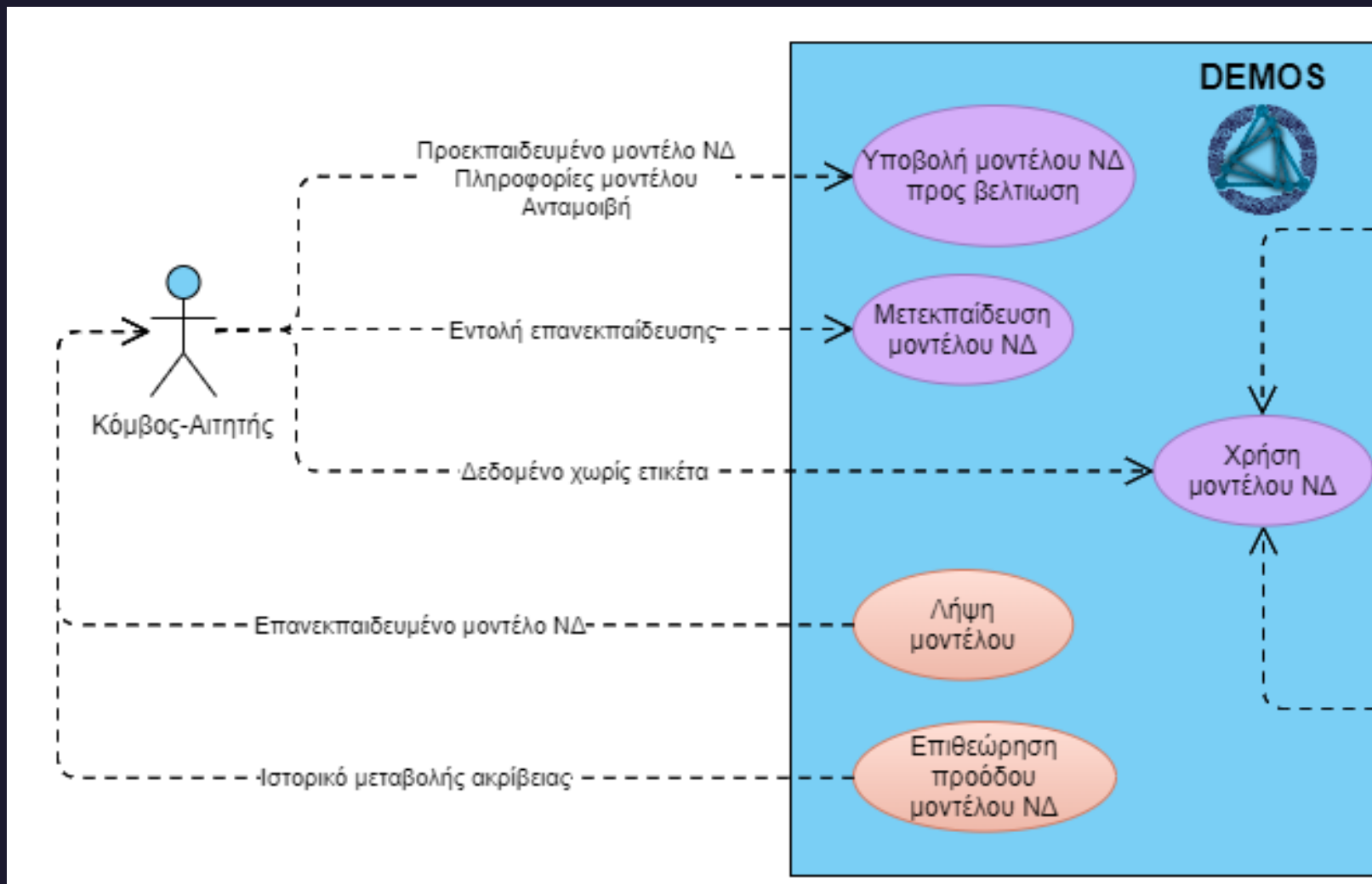
Το μοντέλο επανεκπαιδεύεται με τα νέα δεδομένα και επαναξιολογείται η ακρίβειά του



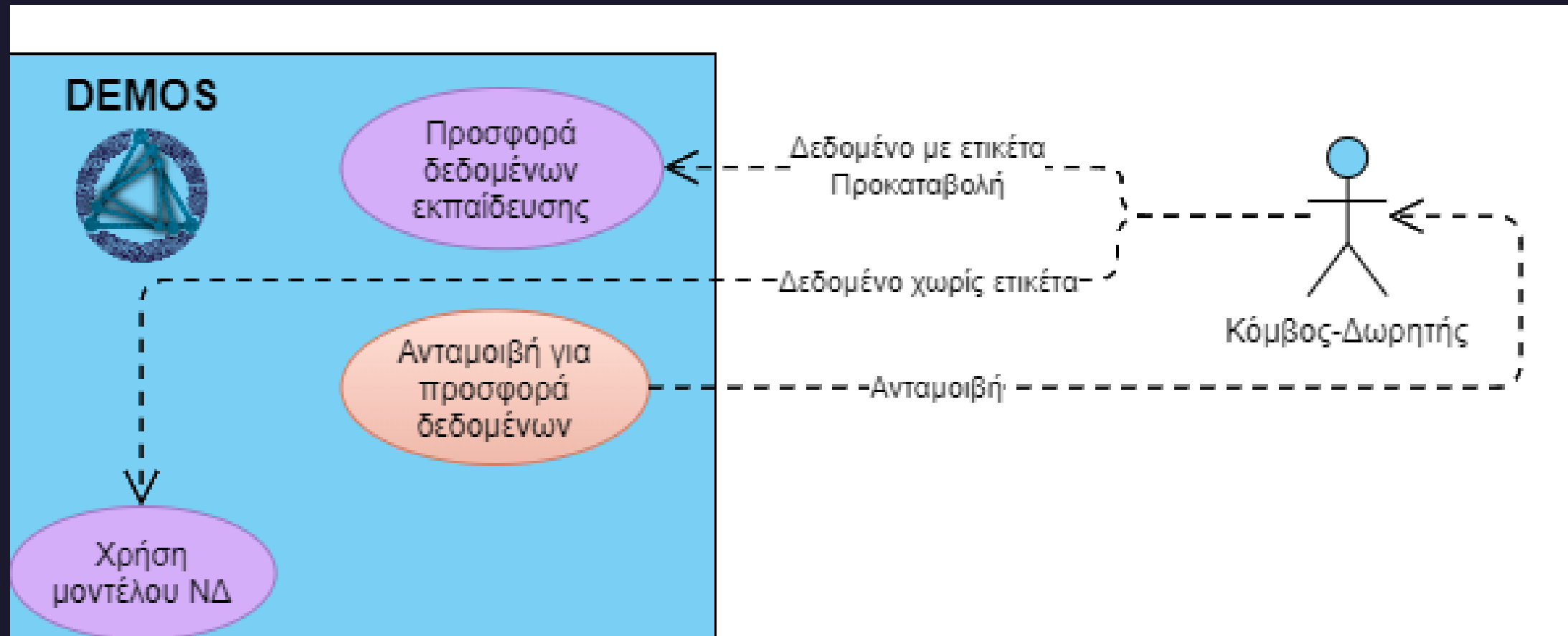
Διάγραμμα τύπων χρηστών



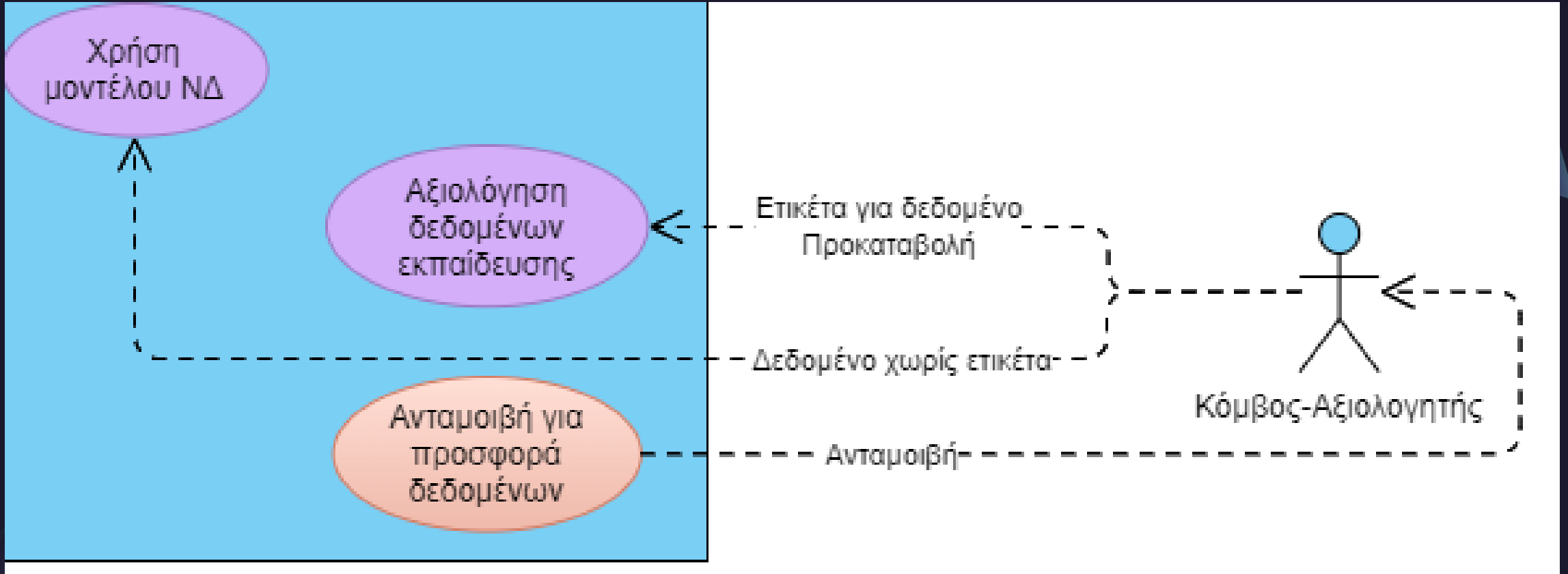
Κόμβος - Αιτητής



Κόμβος - Δωρητής









Κόμβος - Αξιολογητής



Μηχανισμός Κινήτρου

Προσφέρει οικονομικά κίνητρα ώστε να ενθαρρύνει την ορθή χρήση και να τιμωρεί την κακόβουλη ως εξής:

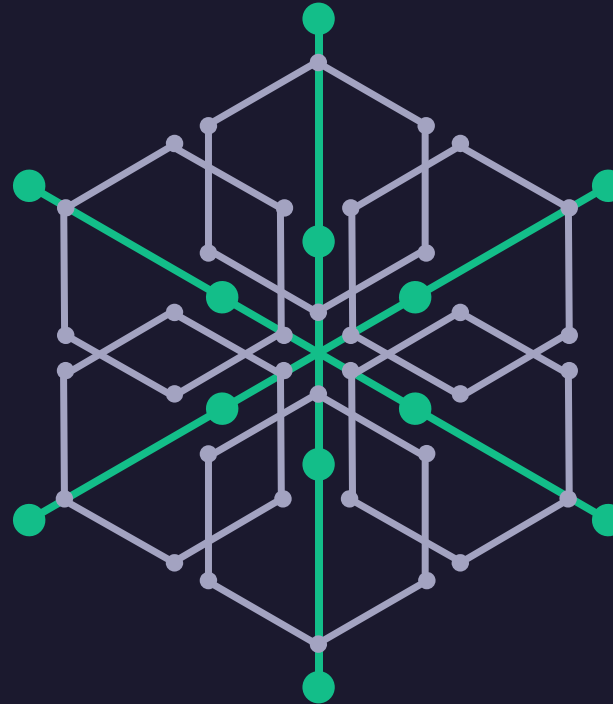
- 1** Ο Κόμβος-Αιτητής θέτει τον αριθμό ψήφων για επικύρωση δεδομένων M 
- 2** Ένας Κόμβος-Δωρητής προσφέρει κάποιο δεδομένο με αντίστοιχη ετικέτα E 
- 3** Οι Κόμβοι-Αξιολογητές ψηφίζουν την ετικέτα που πιστεύουν ότι είναι καταλληλότερη για το δεδομένο (χωρίς να βλέπουν την E) 
- 4** Κάποια πιθανή ετικέτα E' μαζεύει M ψήφους και το δεδομένο θεωρείται πλέον επικυρωμένο 
- 5** Αν $E' = E$ ο κόμβος-δωρητής ανταμείβεται 
- 6** Οι Κόμβοι-Αξιολογητές που ψήφισαν την E' ανταμείβονται 

Μοντέλο Νευρωνικού Δικτύου

Το μοντέλο που χρησιμοποιεί το DEMOS
βασίζεται στο μοντέλο αναπαράστασης
φυσικής γλώσσας BERT

Χρησιμοποιείται για NLP
προβλήματα και συγκεκριμένα
προβλήματα ανάλυσης
συναισθημάτων (sentiment
analysis)

Η αξιολόγηση της επίδοσης των
μοντέλων γίνεται με τη μετρική
Ακρίβειας και ένα
προαποθηκευμένο σύνολο
δεδομένων αξιολόγησης (test
dataset)

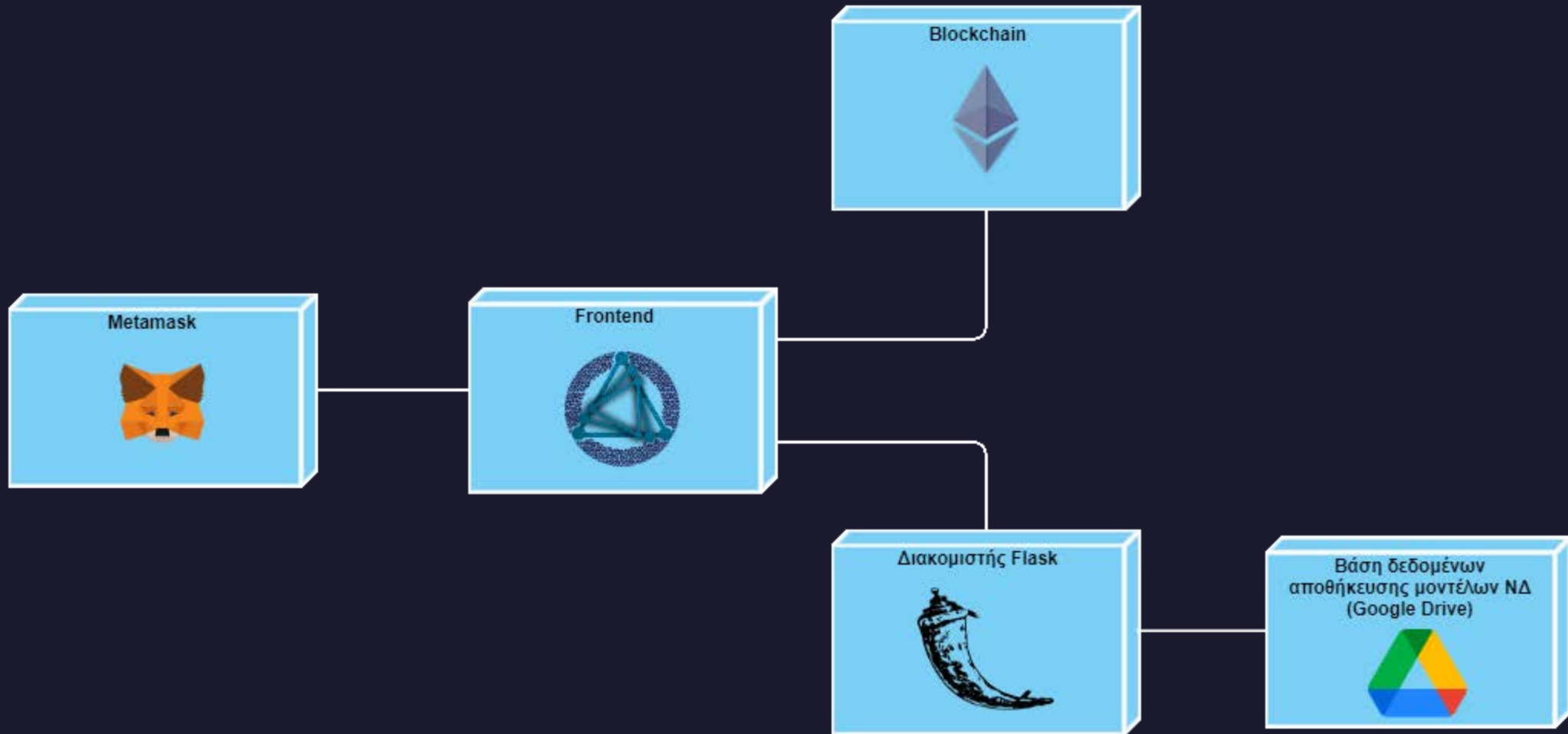


Είναι μοντέλο επιβλεπόμενης
μάθησης επομένως τα δεδομένα
κατηγοριοποιούνται σε κάποιες
προεπιλεγμένες κατηγορίες

Τα δεδομένα εκπαίδευσης είναι
προτάσεις με την αντίστοιχη ετικέτα

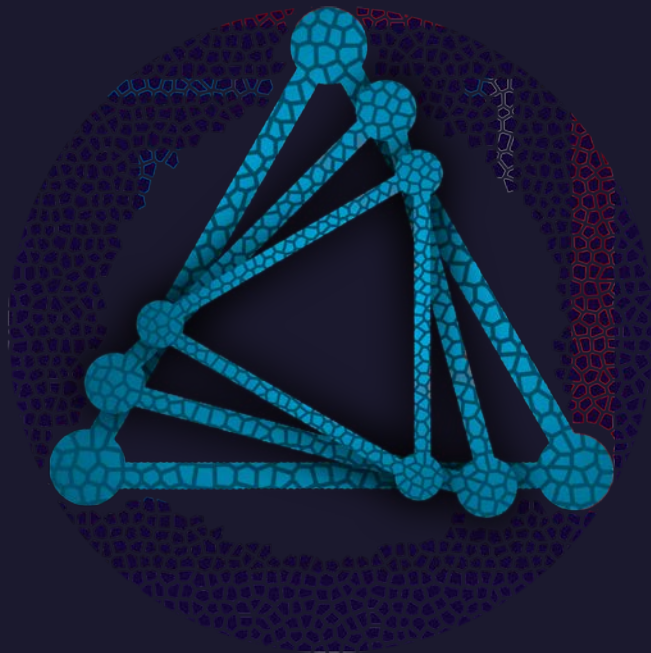
Τα μοντέλα προεκπαιδεύονται πριν
φορτωθούν στο DEMOS

Υλοποίηση



4. Παρουσίαση εφαρμογής

DEMOS



5. Συμπεράσματα

Συμπεράσματα

1

Η συνεχής ανανέωση του συνόλου δεδομένων εκπαίδευσης βελτιώνει την επίδοση

2

Για επανεκπαίδευση με λίγα δεδομένα η μεταβολή στην ακρίβεια είναι μικρή και μπορεί να είναι και αρνητική

3

Το Blockchain ΔΕΝ είναι βάση δεδομένων γι' αυτό απαιτεί πολύ προσεκτικός σχεδιασμός των δομών δεδομένων

4

Στο Blockchain πρέπει να αποθηκεύονται μόνο τα απαραίτητα γιατί τα κόστη συναλλαγών είναι σημαντικά

5

Απαιτείται διεξοδική μελέτη για τον καθορισμό των τιμών προκαταβολής και ανταμοιβής

6

Χρειάζεται μια ισορροπία μεταξύ των υπολογισμών που γίνονται στο Blockchain, αυτών που γίνονται στο frontend και αυτών που γίνονται σε κάποιο εξωτερικό διακομιστή

Πλεονεκτήματα κατανεμημένου συστήματος



1

Η χρονοβόρα και άχαρη συλλογή και αξιολόγηση δεδομένων γίνεται πολύ πιο γρήγορα και έγκυρα

2

Ο Κόμβος-Αιτητής αφιερώνει χρόνο σε πιο εξειδικευμένες διαδικασίες εκπαίδευσης (όπως το fine tuning)

3

Ποικιλομορφία από την κατανεμημένη συλλογή δεδομένων άρα πολύπλευρη εκπαίδευση του μοντέλου

4

Ενδεδεχής ποιοτικός έλεγχος των μοντέλων γιατί γίνεται από πολλούς χρήστες

Πλεονεκτήματα Blockchain

Τα δεδομένα δεν μπορούν να διαγραφούν ή να παραποιηθούν

Δεν απαιτείται σύνδεση σε ηλεκτρονικό περιβάλλον τράπεζας για τις συναλλαγές

Η δημιουργία των smart contracts είναι εύκολη και προσαρμόσιμη στις ανάγκες της εφαρμογής

Το υπόλοιπο των smart contracts λογαριασμών είναι ασφαλές



Κανένας χρήστης δεν αποκλείεται από τη χρήση της εφαρμογής

Οι κόμβοι διατηρούν την ανωνυμία τους

Σας ευχαριστώ για την προσοχή σας

Δημήτρης Κυριάκου

<https://gitlab.com/netmode/blockchain-ml>

