

**ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ**

**ΤΕΛΙΚΗ ΕΡΓΑΣΙΑ**

**ΘΕΜΑ 5:**

**ML-AGENTS(MACHINE LEARNING AGENTS)**

Από τον φοιτητή:

ΙΩΑΝΝΗΣ-ΙΑΣΩΝ ΜΠΟΪΔΑΝΗΣ – Π19217

Καθηγητής: ΘΕΜΗΣ ΠΑΝΑΓΙΩΤΟΠΟΥΛΟΣ

Τηλέφωνο επικοινωνίας: 6971848447

Email: jjboi00113@gmail.com

*Πειραιάς, Ιούνιος 2023*

**ΠΕΡΙΕΧΟΜΕΝΑ**

1.ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ……………………………………3

2. ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΘΕΩΡΗΤΙΚΗΣ ΒΑΣΗΣ……………………………….4

3. ΣΧΕΔΙΑΣΤΙΚΕΣ ΑΠΟΦΑΣΕΙΣ ΚΑΙ ΣΤΟΙΧΕΙΑ ΥΛΟΠΟΙΗΣΗΣ……..7

4. ΠΑΡΑΔΕΙΓΜΑΤΙΚΗ ΕΚΤΕΛΕΣΗ ΚΑΙ ΠΑΡΑΔΕΙΓΜΑΤΑ………….10

5. ΕΓΚΑΤΑΣΤΑΣΗ…………………………………………………………….…..13

6. ΑΝΑΛΥΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΠΡΟΒΛΗΜΑΤΩΝ………………………16

7. ΣΤΑΤΙΣΤΙΚΑ TRAINING………………………………..……………………17

8. ΒΙΒΛΙΟΓΡΑΦΙΑ…………………………………………..……………………20

**Machine learning Agents**

**1.Περιγραφή του προβλήματος**

Κύριος στόχος της εργασίας είναι η αξιοποίηση της τεχνολογίας «πράκτορες μηχανικής μάθησης της Unity» σε μία εφαρμογή. Μέσω αυτής της εργασίας επιτυγχάνεται η εξοικείωση με τέτοιες τεχνικές. Στην δική μου υλοποίηση της εφαρμογής αξιοποιώ και επεμβαίνω στην υπάρχουσα εργασία πρότυπο της Unity που αναφέρεται σε MLAgents όπου μαθαίνουν να παίζουν ποδόσφαιρο.

Συγκεκριμένα μορφοποιώ τους πράκτορες σχεδιαστικά, προσθέτω 2 νέα είδη παιχτών, τον αμυντικό και τον μέσο, αλλάζω τις ικανότητες και δυνατότητες των επιθετικών κι τερματοφυλάκων. Επιπρόσθετα εντάσσω έναν πίνακα σκορ και με βάση αυτόν παρουσιάζω μία νέα παράμετρο στην συνάρτηση ανταμοιβής έτσι ώστε οι πράκτορες να «κοιτάνε» το σκορ, και αν χάνουν με διαφορά 3ων γκολ να θέλουν περισσότερο να βάλλουν γκολ καθώς και θα πάρουν μεγαλύτερο βραβείο. Αντίστοιχα ο αντίπαλος επιθετικός θα θέλει να μένει η διαφορά των γκολ γιατί χάνει λιγότερα βραβεία ο ίδιος.   
Όση ώρα περνά και δεν μπαίνει γκολ ο επιθετικός χάνει ανταμοιβή άρα θέλει να βάλει γκολ όσο το δυνατό πιο σύντομα, αλλά αν χάνει με 3 γκολ διαφορά θα χάνει περισσότερη ανταμοιβή και θα είναι «λυπημένος». Αντίστοιχα οι άλλοι 3ις ρόλοι, όσο δεν τρώνε γκολ είναι «χαρούμενοι» κι παίρνουν ανταμοιβή, αλλά αν χάνουν με παραπάνω από τρία γκολ θα θέλουν να βάλουν και οι ίδιοι γκολ για να πάρουν ανταμοιβή. Να σημειωθεί πως οι MLAgents δεν έχουν πραγματικά συναισθήματα καθώς και δεν είναι affective agents.

**2.Περιγραφή θεωρητικής βάσης**

Το Unity ML-Agents Toolkit είναι ένα ανοικτού κώδικα έργο που μου επέτρεψε να δημιουργήσω ένα προσομοιωμένο περιβάλλον χρησιμοποιώντας το Unity Editor και να αλληλεπιδρώ με αυτό μέσω του Python API. Το εργαλείο παρέχει το ML-Agents SDK που περιλαμβάνει όλη τη λειτουργικότητα που απαιτείται για να φτιαχτεί το περιβάλλον μέσα στο Unity Editor, καθώς και τα βασικά αρχεία κώδικα C# για την κατασκευή της διαδικασίας μάθησης.

Οι τρεις κύριες οντότητες στο ML-Agents SDK είναι οι Αισθητήρες (Sensors), οι Πράκτορες (Agents) και μια Ακαδημία (Academy). Το component του Πράκτορα χρησιμοποιείται για να υποδείξει ότι ένα GameObject εντός μιας σκηνής είναι ένας Πράκτορας και μπορεί να συλλέξει παρατηρήσεις, να πραγματοποιήσει ενέργειες και να λαμβάνει ανταμοιβές. Ο πράκτορας μπορεί να συλλέξει παρατηρήσεις χρησιμοποιώντας διάφορους αισθητήρες που αντιστοιχούν σε διάφορες μορφές πληροφοριών, όπως απεικονίσεις, αποτελέσματα ακτίνων (ray-cast), ή διανύσματα αυθαίρετου μήκους. Κάθε component του Πράκτορα περιέχει μια πολιτική που επισημαίνεται με ένα όνομα συμπεριφοράς.

Οποιοσδήποτε αριθμός πρακτόρων μπορεί να έχει μια πολιτική με το ίδιο όνομα συμπεριφοράς. Αυτοί οι πράκτορες θα εκτελέσουν την ίδια πολιτική και θα μοιραστούν δεδομένα εμπειρίας κατά τη διάρκεια της εκπαίδευσης. Επιπλέον, μπορεί να υπάρχουν οποιοδήποτε αριθμός ονομάτων συμπεριφοράς για πολιτικές μέσα σε μια σκηνή, επιτρέποντας την απλή δημιουργία πολλαπλών πρακτόρων που εκτελούν διαφορετικούς τύπους συμπεριφοράς. Μια πολιτική μπορεί να αναφέρεται σε διάφορους μηχανισμούς λήψης αποφάσεων, συμπεριλαμβανομένων των εισόδων του παίκτη, σεναρίων που έχουν καθοριστεί εξωτερικά, ενσωματωμένων μοντέλων νευρωνικών δικτύων ή μέσω αλληλεπίδρασης μέσω του API της Python. Οι πράκτορες μπορούν να ζητήσουν αποφάσεις από τις πολιτικές είτε σε σταθερό είτε σε δυναμικό διάστημα, όπως καθορίζεται από τον προγραμματιστή του περιβάλλοντος.

Η συνάρτηση ανταμοιβής, που χρησιμοποιείται για να παρέχει ένα σήμα μάθησης στον πράκτορα. Επίσης, η προσομοίωση μπορεί να τεθεί σε κατάσταση "τέλους" είτε σε επίπεδο ατομικού πράκτορα είτε σε επίπεδο του περιβάλλοντος ως σύνολο. Αυτό συμβαίνει είτε μέσω κλήσης σε ένα σενάριο του Unity είτε όταν επιτευχθεί ένα προκαθορισμένο μέγιστο αριθμό βημάτων.

Η Ακαδημία είναι μια μοναδική οντότητα μέσα στην προσομοίωση και χρησιμοποιείται για να καταγράφει τα βήματα της προσομοίωσης και να διαχειρίζεται τους πράκτορες. Η Ακαδημία περιέχει επίσης τη δυνατότητα ορισμού παραμέτρων περιβάλλοντος, οι οποίες μπορούν να χρησιμοποιηθούν για να αλλάξουν τη διαμόρφωση του περιβάλλοντος κατά την εκτέλεση. Ειδικότερα, ορισμένες πτυχές της φυσικής του περιβάλλοντος και των απεικονίσεων, των μεγεθών και της ύπαρξης των GameObjects ελέγχονται μέσω εκθέσιμων παραμέτρων που μπορούν να ανακληθούν και να αλλοιωθούν κατά τη διάρκεια της εκπαίδευσης. Για παράδειγμα, η βαρύτητα στο περιβάλλον μπορεί να διακυμαίνεται σε κάθε σταθερό διάστημα ή μπορούν να προστεθούν επιπλέον εμπόδια όταν ένας πράκτορας φτάνει σε ένα συγκεκριμένο επίπεδο επάρκειας. Αυτό δίνει τη δυνατότητα αξιολόγησης ενός πράκτορα σε μια διαίρεση εκπαίδευσης/δοκιμής μεταβολών του περιβάλλοντος και ευνοεί τη δημιουργία σεναρίων μάθησης με βάση το πρόγραμμα σπουδών (curriculum learning scenarios).

Κάθε πράκτορας στην συγκεκριμένη υλοποίηση μπορεί να κινηθεί μπροστά πίσω δεξιά και αριστερά, καθώς και να κάνει δεξιά ή αριστερή περιστροφή. Οι πράκτορες έχουν επίσης όραση χρησιμοποιώντας ray cast, ανάλογα το τι βλέπουν, έχουν μία ιεράρχηση στην σημαντικότητα των πραγμάτων που διακρίνουν, συγκεκριμένα έχουν προτεραιότητα στο να επικεντρώνονται στην μπάλα και στα τέρματα αντί στους παίχτες.  
Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό, λογισμικό πολυμέσων

Περιγραφή που δημιουργήθηκε αυτόματα

**2.1.Δομές και αλγόριθμοι**

Οι λειτουργίες της εργαλειοθήκης και η εκπαίδευση των ευφυή πρακτόρων για την δική μου υλοποιήση πραγματοποιείται από τον αλγόριθμο ενισχυτικής μάθησης Proximal Policy Optimization (PPO) και αλγόριθμους Heuristic για την συμπεριφορά τους. Όμως κατά την διάρκεια του training δεν χρησιμοποιήθηκαν Heuristic συμπεριφορές.

**2.2. Αναπαράσταση γνώσης**

Όσο αναφορά την αναπαράσταση γνώσης, από την εργαλειοθήκη της Unity προϋπήρχε νευρωνικό δίκτυο για τους τερματοφύλακες και τους επιθετικούς. Καθώς και η εφαρμογή μου είχε μέσους και αμυντικούς, έπρεπε να γίνει περεταίρω εκπαίδευση. Η οποία έχει αποθηκευτεί ως μοντέλο νευρωνικού δικτύου.

**3.Σχεδιαστικές αποφάσεις**Για έξοδο από την εφαρμογή πατήστε esc

Όπως αναφέρθηκε και στην περιγραφή του προβλήματος ήθελα να προσθέσω 2 νέους παίχτες ποδοσφαίρου, δηλαδή να δώσω στους πράκτορες νέα Positions.

Εικόνα που περιέχει κείμενο, γραμματοσειρά, στιγμιότυπο οθόνης, σχεδίαση

Περιγραφή που δημιουργήθηκε αυτόματα Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό, λογισμικό πολυμέσων

Περιγραφή που δημιουργήθηκε αυτόματα

Μορφοποίησα τις δεξιότητες και τις δυνατότητες τους ως εξής:

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

Περιγραφή που δημιουργήθηκε αυτόματα

Ήθελα να έχουν πιο αντιπροσωπευτικά χαρακτηριστικά ανάλογα την θέση στην οποία παίζουν.

Μορφοποίησα τα 3D μοντέλα των χαρακτήρων για πιο τυχαίο σουταρισμα της μπάλας:

Εικόνα που περιέχει στιγμιότυπο οθόνης, καρτούν

Περιγραφή που δημιουργήθηκε αυτόματαΕικόνα που περιέχει στιγμιότυπο οθόνης, καρτούν, κύκλος

Περιγραφή που δημιουργήθηκε αυτόματα

Έφτιαξα έναν λειτουργικό πίνακα σκορ όπου αυξάνει τα γκολ της κάθε ομάδας.

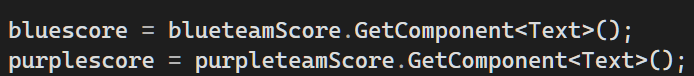
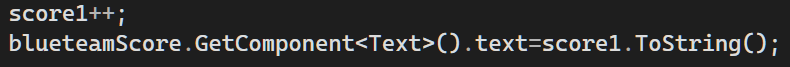
Εικόνα που περιέχει στιγμιότυπο οθόνης, κείμενο, σχεδίαση

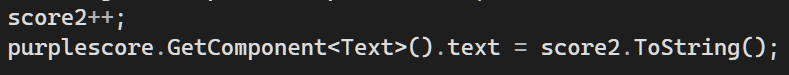
Περιγραφή που δημιουργήθηκε αυτόματα



Εικόνα που περιέχει κείμενο, γραμματοσειρά, στιγμιότυπο οθόνης, γραμμή

Περιγραφή που δημιουργήθηκε αυτόματα



Η συνάρτηση ανταμοιβής επιβραβεύει τον αμυντικό τον μέσο και τον τερματοφύλακα για κάθε βήμα που κάνουν ενώ δεν έχουν φάει γκολ. Αντίθετα ο επιθετικός χάνει βραβεία για κάθε βήμα που κάνει και δεν έχει βάλει γκολ. Ανάλογα το σκορ, συγκεκριμένα αν μία από τις δύο ομάδες χάνει με 3 γκολ διαφορά, τότε όλοι οι πράκτορες της θα χάνουν βραβεία πιο γρήγορα κάθε βήμα που κάνουν. Συνεπώς θα θέλουν να βάλουν γκολ περισσότερο, ή να μην αφήσουν το σκορ να κατρακυλήσει πιο πολύ.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

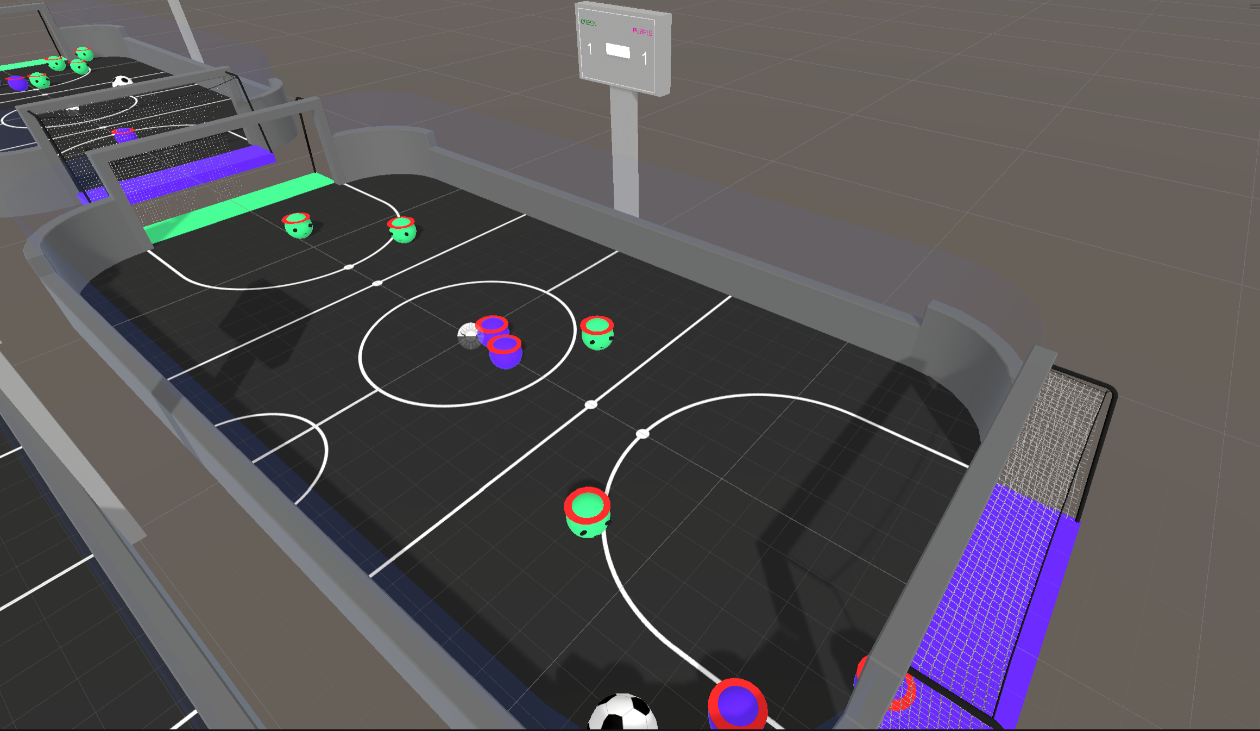
Περιγραφή που δημιουργήθηκε αυτόματα

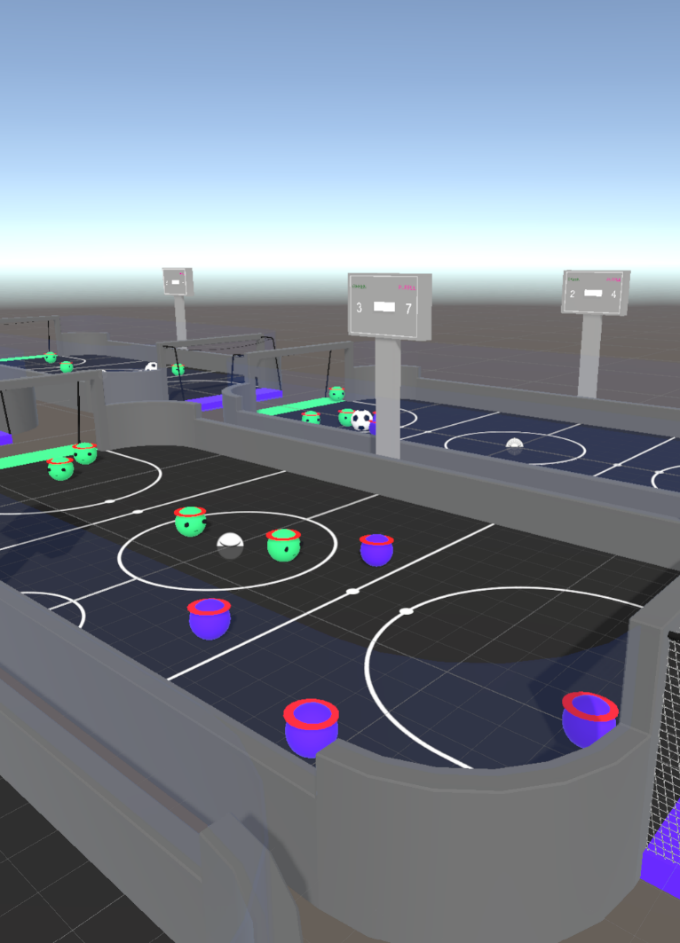
O μέγιστος αριθμών βημάτων που μπορεί να κάνει κάποιος πράκτορας είναι 5000 αν δεν εκπληρώσει τον στόχο του  
Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό, λογισμικό πολυμέσων

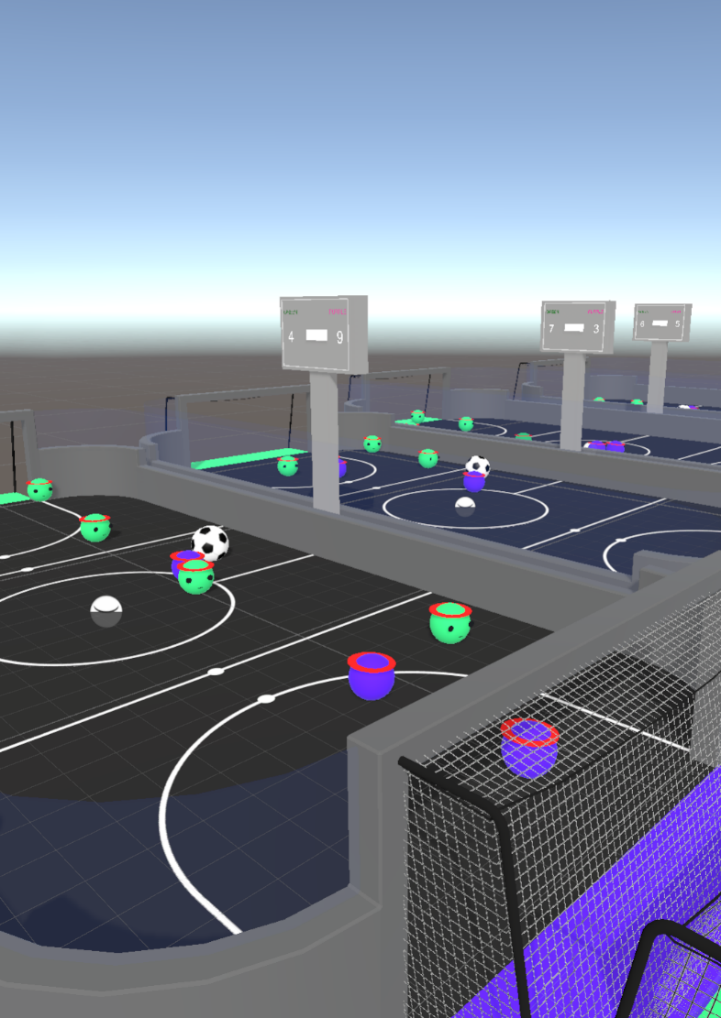
Περιγραφή που δημιουργήθηκε αυτόματα

**4.ΠΑΡΑΔΕΙΓΜΑΤΙΚΗ ΕΚΤΕΛΕΣΗ**

Στα παρακάτω στιγμιότυπα οθόνης βλέπουμε τους πράκτορες να παίζουν ποδόσφαιρο και να αυξάνεται το σκορ.







Και από την διαδικασία του training:  
Εικόνα που περιέχει στιγμιότυπο οθόνης, λογισμικό πολυμέσων, λογισμικό γραφικών, παιχνίδι

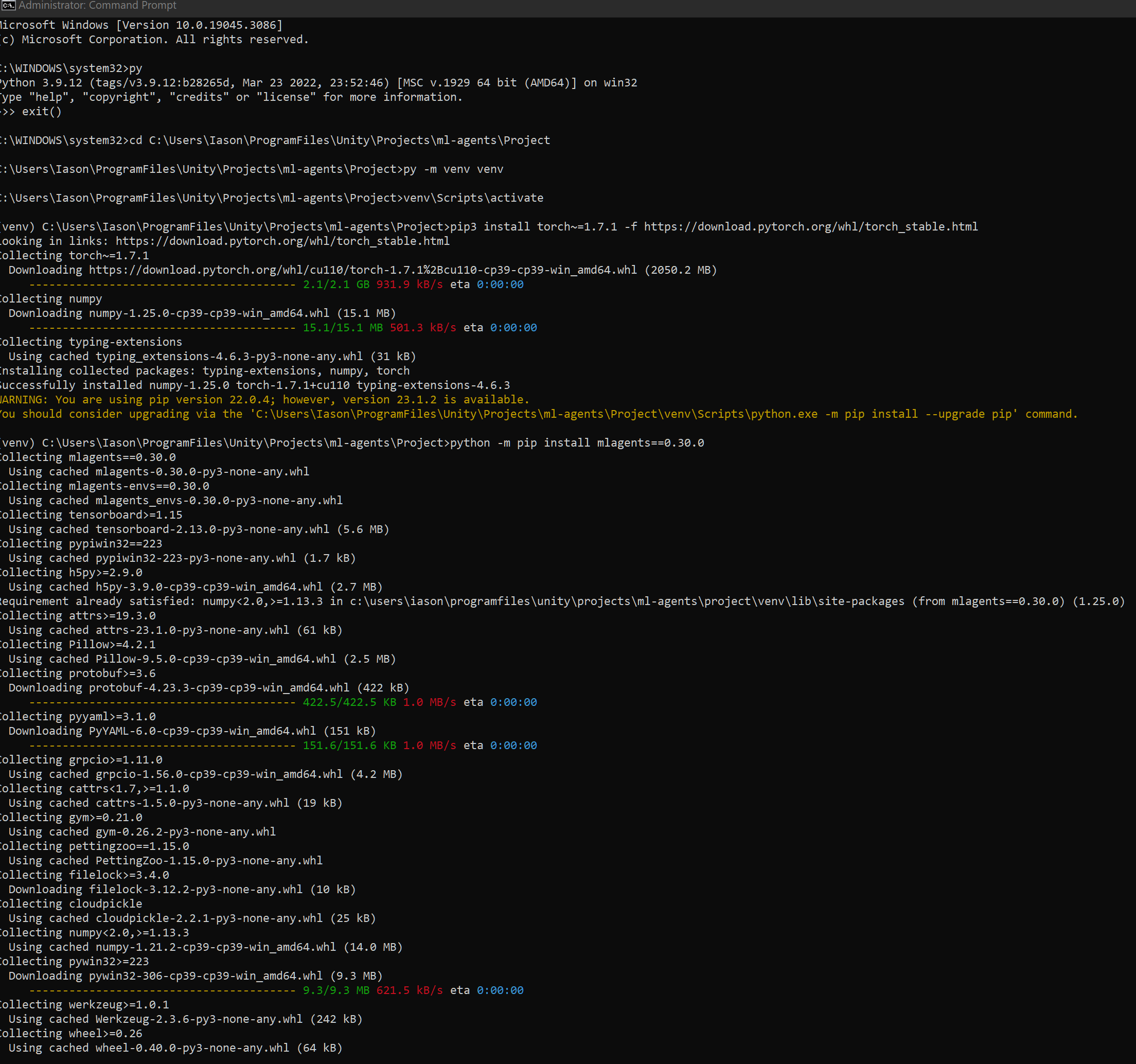
Περιγραφή που δημιουργήθηκε αυτόματα

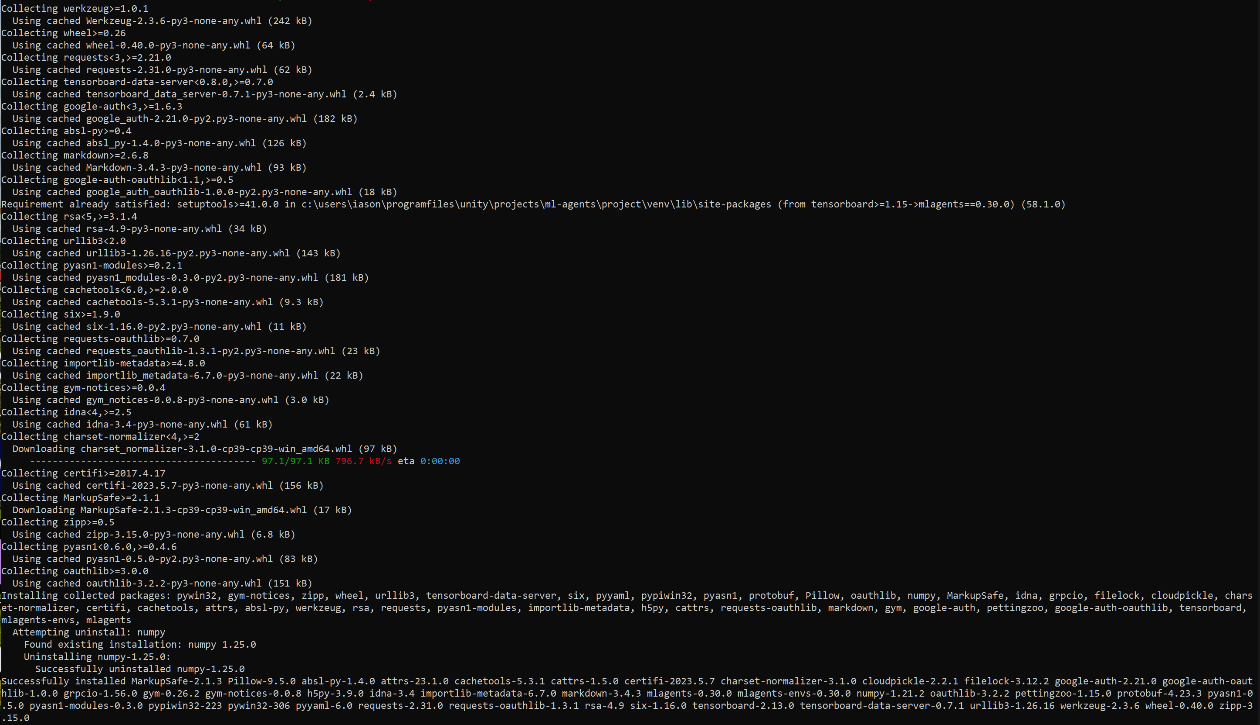
Εικόνα που περιέχει στιγμιότυπο οθόνης, παιχνίδι

Περιγραφή που δημιουργήθηκε αυτόματα

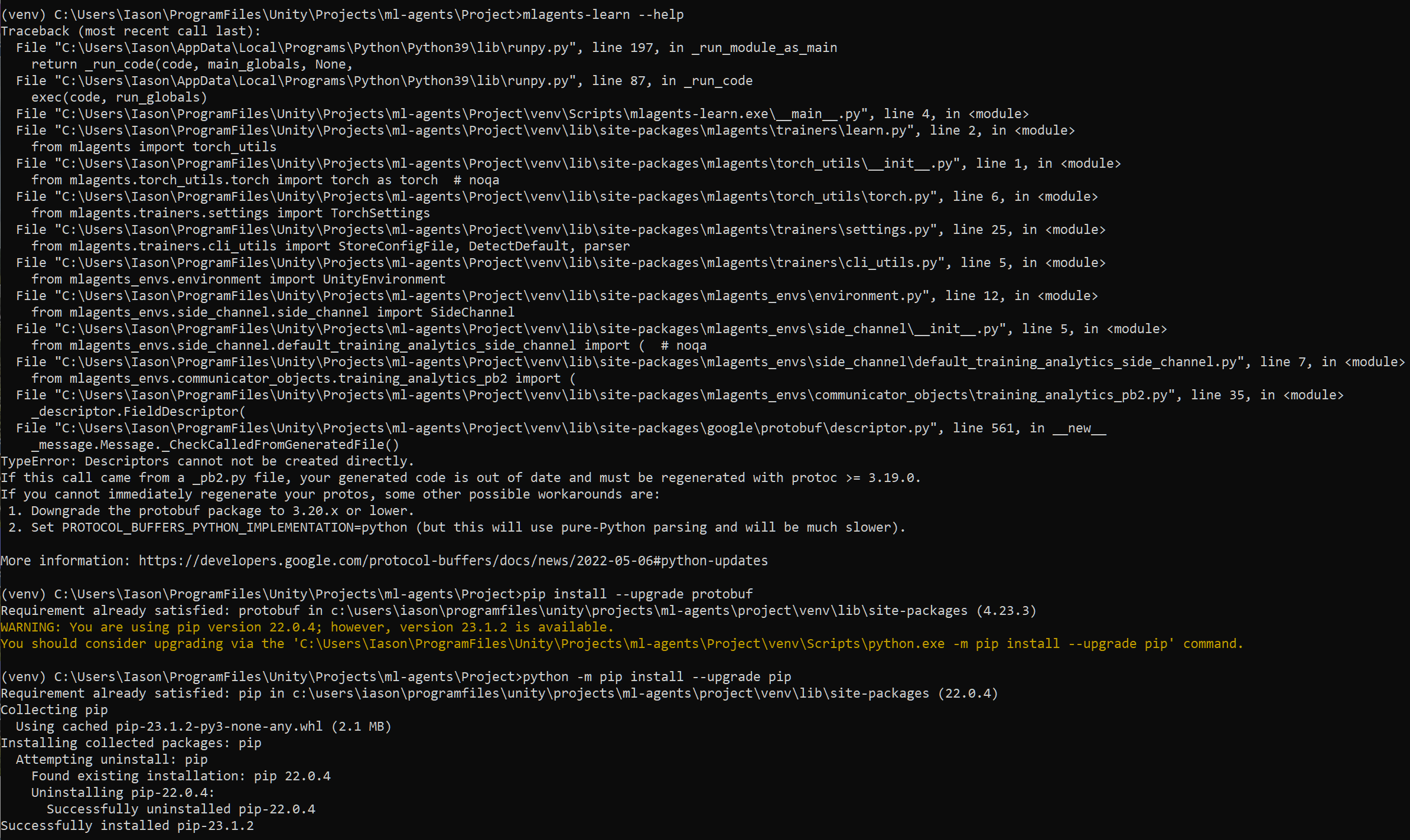
**5. Εγκατάσταση**

Αρχικά αναβάθμισα την έκδοση της python μου για να είναι συμβατή με τις οδηγίες του επίσημου repository της Unity. Έφτιαξα φάκελο για να βάλω το εικονικό μου περιβάλλον μέσα, το ενεργοποίησα και στην συνέχεια κατέβασα τα απαραίτητα εργαλεία.

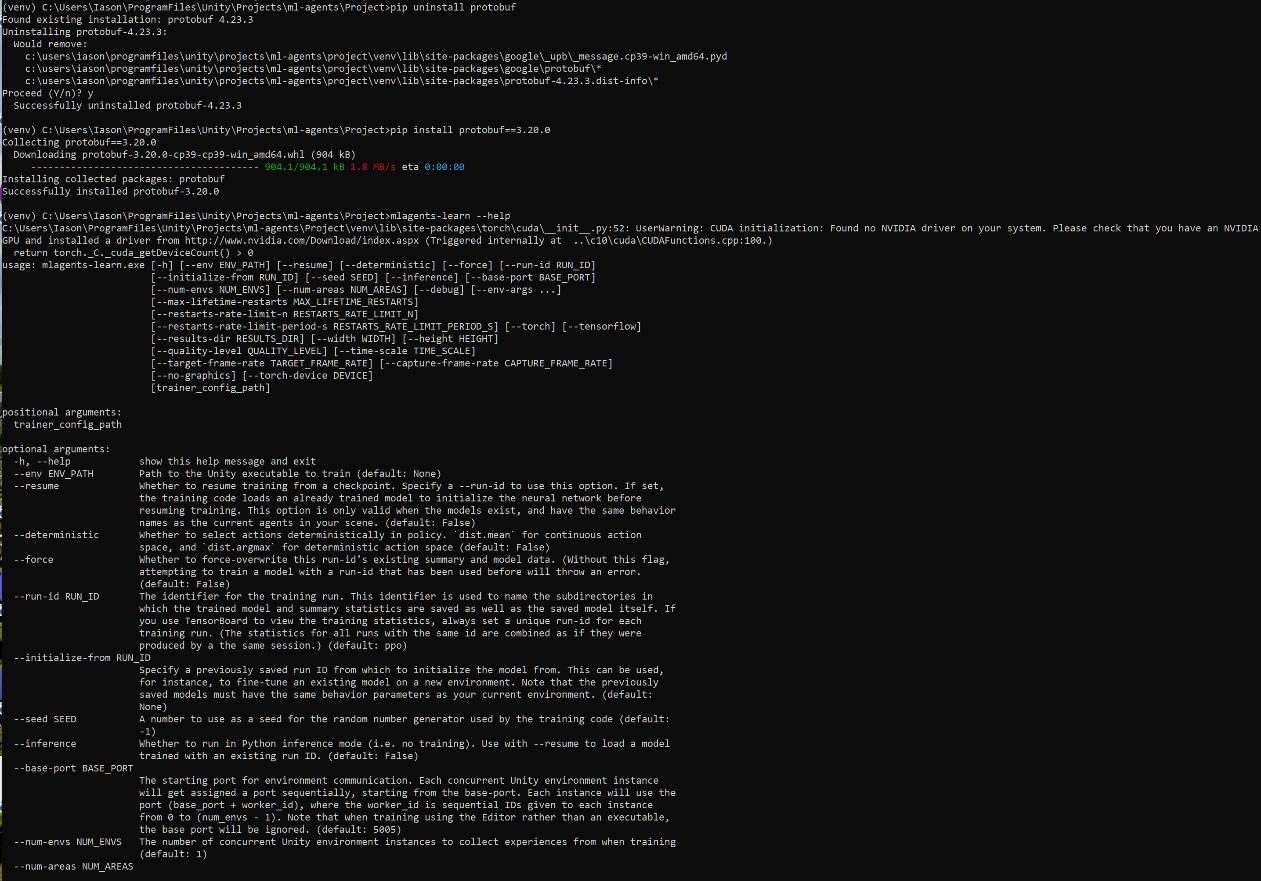




Αφού κατέβηκαν τέσταρα αν έχουν κατέβει σωστά.



Επειδή είχα κάποια ερορ, χρειάστηκε να κατεβάσω ακόμα ένα πακέτο, και εν τέλη έτρεξε το toolkit των mlagents



Και βλέπουμε ότι η εντολή εκτελείται χωρίς κάποιο θέμα, οπότε ολοκληρώθηκε η διαδικασία της εγκατάστασης.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

Περιγραφή που δημιουργήθηκε αυτόματα

Και επίσης μέσα στο προτζεκτ κατέβασα και το αντίστοιχο πακέτο για τους πράκτορες.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

Περιγραφή που δημιουργήθηκε αυτόματα

**6. Αναλυτική περιγραφή προβλημάτων**

Μερικά θέματα στην εκπαίδευση των πρακτόρων ήταν τα εξής:

* Πρόβλημα στην εκπαίδευση πολλών ομάδων μαζί
* Δεν χρησιμοποιώ multy-agent trainer

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, λογισμικό, λογισμικό πολυμέσων

Περιγραφή που δημιουργήθηκε αυτόματα

* Οι πράκτορες στις θέσεις άμυνα κέντρο και τέρμα «τρώνε» κάποιες φορές επίτηδες γκολ για να χάνουν με 3 γκολ διαφορά, έτσι ώστε να πάρουν περισσότερα βραβεία όταν μειώσουν την διαφορά βάζοντας γκολ, πλην του επιθετικού που επιχειρεί συνέχεια να βάλει γκολ.
* Μερικές φορές ο τερματοφύλακας αφήνει κενή την εστία του γιατί προσηλώνεται στην κατεύθυνση της μπάλας
* Ο αμυντικός παίζει και σαν τερματοφύλακας όταν γίνεται αυτό
* Κανένας πράκτορας δεν τα πάει τόσο καλά με τις «σπόντες» της μπάλας
* Μερικές φορές αν χάνουν πολύ ώρα με μεγάλη διαφορά στο σκορ τότε δυσκολεύονται πολύ να βάλουν γκολ γιατί έχουν χάσει πολλά βραβεία, και συνήθως δεν μπορεί να «γυρίσει» το ματς σε σχέση αν μείνει κοντά ο αγώνας

Εικόνα που περιέχει στιγμιότυπο οθόνης

Περιγραφή που δημιουργήθηκε αυτόματα

**7. Στατιστικά Training**

**Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

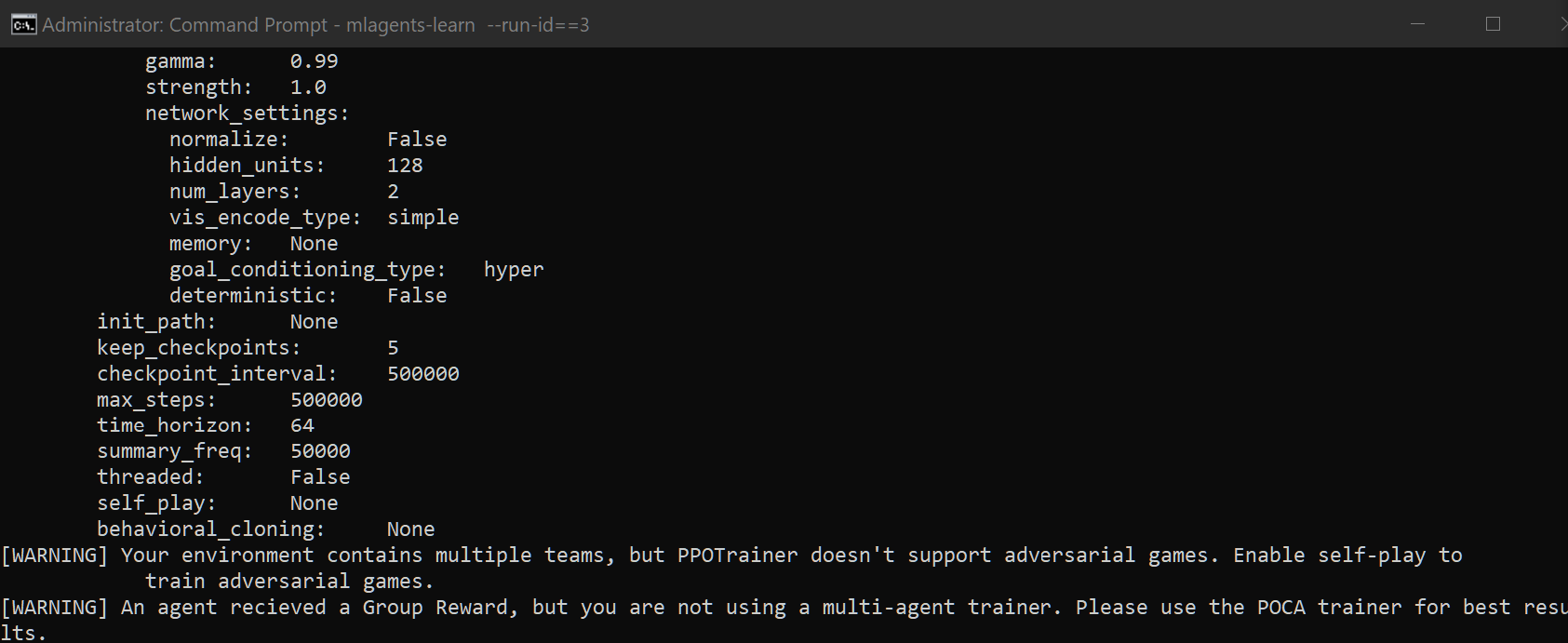
Περιγραφή που δημιουργήθηκε αυτόματα**

Τα αρχεία του training μου βρίσκονται στον φάκελο results

Για να ξεκινήσει η διαδικασία της εκπαίδευσης τρέχω την παρακάτω εντολή:

****

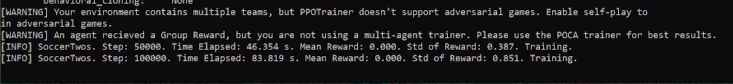
Πατάμε play στο unity.

****

Και μετά τα warnings που ανέφερα πιο πριν ξεκινάνε οι μετρήσεις.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά

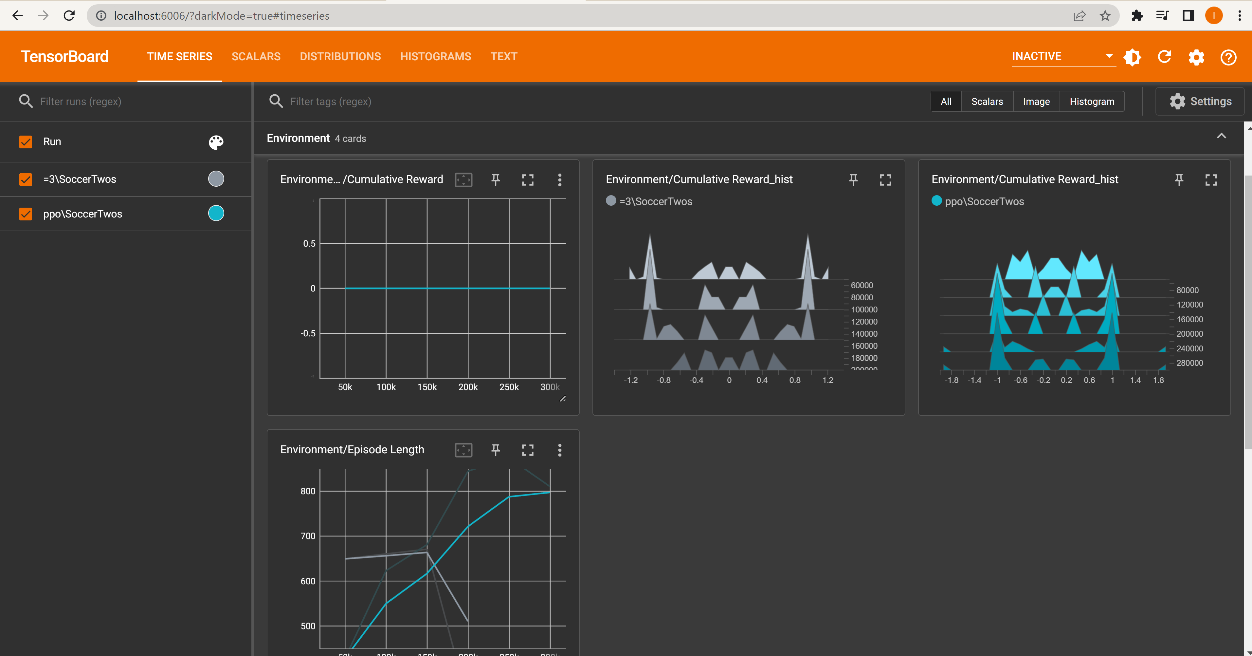
Περιγραφή που δημιουργήθηκε αυτόματα

****

Όπως βλέπουμε δημιουργήθηκαν δύο brain(neural networks)

Για κάθε ομάδα, και εμφανίζονται τα μέγιστα βήματα που μπορούν να κάνουν οι πράκτορες, συνεπώς ξεκίνησε νέα διαδικασία εκπαίδευσης, η μέση ανταμοιβή είναι μηδέν αφού ο επιθετικός χάνει και οι υπόλοιποι κερδίζουν βραβεία καθώς δεν σκοράρεται τέρμα.

Στα στατιστικά της εκπαίδευσης τα βλέπουμε αν μεταβούμε στην ανοιχτή θύρα 6006. Τα οποία αναγράφουν την μέση ανταμοιβή που αναφέρθηκε πριν, το μέσο μήκος σε χρόνο της εκπαίδευσης το οποίο αυξάνεται καθώς οι πράκτορες μαθαίνουν να κερδίζουν ανταμοιβές( να βάζουν γκολ) και έτσι να αυξάνουν τα διαθέσιμα βήματα που μπορούν να κάνουν.

****

και η απώλειες της εκπαίδευσης.

****

**8. Βιβλιογραφία**

[**https://arxiv.org/pdf/1809.02627.pdf**](https://arxiv.org/pdf/1809.02627.pdf)

[**https://github.com/Unity-Technologies/ml-agents**](https://github.com/Unity-Technologies/ml-agents)

[**https://www.youtube.com/watch?v=zPFU30tbyKs&ab\_channel=CodeMonkey**](https://www.youtube.com/watch?v=zPFU30tbyKs&ab_channel=CodeMonkey)