בהחלט. זו שאלה מצוינת שמסכמת את כל שלב התכנון ועוברת לשלב הביצוע. אחרי שתהיה לך תשובה ברורה לשאלה הזו, יהיה לך "ספר בישול" מדויק לאימון המערכת.

לפי אסטרטגיית "הפרד ומשול" שגיבשנו, אנחנו צריכים לאמן **שלושה מודלים עיקריים** באופן נפרד. כל אימון כזה יתבצע במחברת Google Colab Pro+ ייעודית.

להלן פירוט מלא עבור כל מודל.

### **1. מודל מנוע זיהוי מצבי השוק (Regime Detection Engine - RDE)**

* **המודל:** זוהי הרשת ההיברידית **Transformer + VAE** שלומדת באופן בלתי-מפוקח את "מפת השוק" ומייצרת את ה-Regime\_Vector.
* **הערכת זמן אימון:** **2 עד 8 שעות.**
  + **תלוי ב:** גודל הטרנספורמר, מספר האיטרציות (epochs) שיידרשו להתכנסות, וגודל ה-dataset. מכיוון שהאימון אינו כולל סימולציית מסחר מורכבת, הוא צפוי להיות המהיר ביותר מבין השלושה.
* **קריטריונים לאימון:**

| **קריטריון** | **פירוט** |
| --- | --- |
| **מטרת האימון** | ללמד את המודל לדחוס את וקטורי מאפייני ה-MMD לייצוג רציף, עשיר ובעל משמעות במרחב סמוי (latent space). |
| **סוג הלמידה** | **למידה בלתי-מפוקחת (Unsupervised Learning)**. אין צורך בתוויות. |
| **סביבת האימון** | מחברת Colab שבה נטען את סט הנתונים המלא של וקטורי ה-MMD\_Feature\_Vector ההיסטוריים. |
| **פונקציית הפסד (Loss)** | הפסד משולב של VAE: **Reconstruction Loss + (beta \* KL Divergence Loss)**. המטרה היא למזער את ההפסד הזה. |
| **תנאי עצירה** | האימון יסתיים כאשר ההפסד על סט הולידציה (Validation Set) מפסיק להשתפר לאורך מספר איטרציות. |
| **התוצר הסופי** | קובץ מודל מאומן אחד: **hybrid\_regime\_engine.pth**. |

### **2. מודל מנוע ניהול הסיכונים (Risk Management Sub-system - M-RMS)**

* **המודל:** זהו סוכן ה-**MAPPO** הייעודי שלומד להמליץ על תוכנית ניהול עסקה אופטימלית.
* **הערכת זמן אימון:** **8 עד 24 שעות.**
  + **תלוי ב:** מספר הסינרגיות ההיסטוריות שנמצאו (כמות ה"קרבות" שהוא יכול ללמוד מהם), מורכבות סביבת הסימולציה, וקצב הלמידה של הסוכן. זהו תהליך איטי יותר כי הוא כולל סימולציה של כל עסקה.
* **קריטריונים לאימון:**

| **קריטריון** | **פירוט** |
| --- | --- |
| **מטרת האימון** | ללמד את הסוכן מדיניות שתמקסם את **יחס הסורטינו** של החשבון לאורך זמן, תוך **עמידה מוחלטת** בכללי החשבון. |
| **סוג הלמידה** | **למידת חיזוק (Reinforcement Learning)** באמצעות אלגוריתם **MAPPO**. |
| **סביבת האימון** | סימולטור ייעודי ("מאמן קרב") שעושה את הפעולות הבאות בלולאה: 1. מוצא סינרגיה היסטורית. 2. מבקש מהסוכן תוכנית סיכון. 3. מדמה את תוצאת העסקה. 4. מחשב ומחזיר לסוכן את התגמול. |
| **פונקציית תגמול (Reward)** | הפונקציה המורכבת שהגדרנו, המבוססת על **השינוי ביחס הסורטינו** וכוללת **עונש כבד** על הפרת כללים. |
| **תנאי עצירה** | האימון יסתיים כאשר התגמול הממוצע לעסקה (או יחס הסורטינו הכולל) מפסיק להשתפר ומגיע לרמה יציבה וגבוהה. |
| **התוצר הסופי** | קובץ מודל מאומן אחד: **risk\_manager\_agent.pth**. |

### **3. מודל ליבת ה-MARL הראשית (Main MARL Core / DecisionGate)**

* **המודל:** זהו סוכן ה-**MAPPO** הראשי והמרכזי, בעל המדיניות המשותפת, שלומד לקבל את ההחלטה הסופית EXECUTE / REJECT.
* **הערכת זמן אימון:** **12 עד 36 שעות.**
  + **תלוי ב:** זהו האימון המורכב ביותר, שכן הוא מדמה את כל זרימת המערכת, כולל קריאות למומחים המאומנים. קצב ההתקדמות שלו תלוי במספר הזדמנויות המסחר הכולל שנמצאו בהיסטוריה.
* **קריטריונים לאימון:**

| **קריטריון** | **פירוט** |
| --- | --- |
| **מטרת האימון** | ללמד את הסוכן מדיניות שתדע מתי **לסמוך על המומחים שלה** ומתי לא, כדי למקסם את הרווחיות הכוללת של המערכת. |
| **סוג הלמידה** | **למידת חיזוק (Reinforcement Learning)** באמצעות אלגוריתם **MAPPO**. |
| **סביבת האימון** | הסימולטור המלא של המערכת. הוא מריץ את כל הלוגיקה, **טוען את המודלים המאומנים וה"קפואים"** של ה-RDE וה-M-RMS, ומאפשר לסוכן הראשי ללמוד מהאינטראקציה איתם. |
| **פונקציית תגמול (Reward)** | תגמול פשוט יותר, המבוסס על התוצאה הסופית: +1 על עסקה רווחית, -1 על עסקה מפסידה. ניתן להוסיף בונוס קטן על דחיית עסקה שהייתה יכולה להיות מפסידה. |
| **תנאי עצירה** | האימון יסתיים כאשר מדד הביצועים הכולל של המערכת (למשל, יחס שארפ או רווח כולל על סט ולידציה) מפסיק להשתפר. |
| **התוצר הסופי** | קובץ מודל מאומן אחד: **main\_decision\_gate.pth**. |

לאחר ששלושת המודלים הללו מאומנים ומוכנים, המערכת שלך שלמה וניתן להתחיל בשלב הבדיקות הסופיות והפריסה.