# **תכנון ומימוש מתקדם של מערכת מסחר אלגוריתמית מרובת סוכנים: שילוב סוכנים הטרוגניים, ארכיטקטורות פריסה רובסטיות וטכניקות MARL פיננסיות ייעודיות**

## **חלק I: אסטרטגיית מסחר יסודית וארכיטקטורת נתונים**

### **סעיף 1.1: רכיבי ליבה של אסטרטגיית המסחר האלגוריתמית**

חלק זה מפרק את אבני הבניין היסודיות של אסטרטגיית המסחר המוצעת, תוך פירוט כל אינדיקטור, פרמטרי הפעולה שלו, גישת הניתוח הרב-ממדית (multi-timeframe), והתפיסה החדשנית של "צורות הרמוניות" הנגזרות מצירופי אותות. הבנת מרכיבים אלו חיונית טרם תכנון מערכות הבינה המלאכותית אשר ימכנו וימטבו את ביצוען.

#### **1.1.1. מסגרת מרובת אינדיקטורים: ניתוח MLMI, NW-RQK, FVG ו-LVN**

ליבת האסטרטגיה נשענת על מפגש אותות ממגוון אינדיקטורים, כאשר כל אחד מהם מספק פרספקטיבה ייחודית על דינמיקת השוק.

* **MLMI (Machine Learning Momentum Index):**
  + אינדיקטור זה פועל על גרפים של 30 דקות של נרות הייקן אשי.1 כפי שמפורט ב-1, הוא משלב עקרונות מומנטום מסורתיים (המבוססים על RSI וממוצעים נעים משוקללים - WMA) עם אלגוריתם למידת מכונה מסוג k-Nearest Neighbors (k-NN) במטרה לייצר תחזיות מומנטום אדפטיביות. אות המסחר העיקרי נגזר בעת חצייה של קו התחזית של MLMI עם ה-WMA שלו.1
  + העדפת משתמש מרכזית היא שימוש בפרמטרי ברירת מחדל עבור MLMI (למשל, 'Prediction Data (k)', 'Trend length') לשמירה על רובסטיות בתנאי שוק משתנים.1 משמעות הדבר היא שסוכן ה-MARL, ולא האינדיקטור עצמו, יהיה אחראי על התאמה להקשרי שוק משתנים, בעיקר על ידי מינוף קלט משטר השוק ממנוע ה-MMD.1
* **NW-RQK (Nadaraya-Watson Rational Quadratic Kernel Regression):**
  + פועל גם הוא על גרפים של 30 דקות של נרות הייקן אשי.1 אינדיקטור זה משתמש ברגרסיית ליבה לא-פרמטרית עם ליבת Rational Quadratic. ליבה זו מקנה משקל גבוה יותר לנתונים עדכניים, ומאפשרת תגובה מהירה יותר לשינויי מחירים ויוצרת עקומת מחיר "מותאמת מיטבית".1
  + אותות מסחר נוצרים כאשר עקומת רגרסיה זו משנה את כיוונה, ומצביעה על מעבר לסנטימנט שורי או דובי.1 תכונה קריטית, המודגשת ב-1, היא היותו "non-repainting" (אינו מצייר מחדש ערכים היסטוריים), תכונה הכרחית לאמינות תוצאות בדיקות לאחור (backtesting) ולביצועים עקביים במסחר חי. גם כאן, מועדפים פרמטרי ברירת מחדל (למשל, 'Bandwidth', 'Alpha').1
* **FVG (Fair Value Gap) - LuxAlgo:**
  + אינדיקטור זה מזהה אזורים של חוסר איזון במחיר ופועל על גרפים של 5 דקות של נרות יפניים סטנדרטיים (שינוי ממסגרת הזמן של דקה באסטרטגיה המקורית 1, לפי 1). האות המרכזי לאסטרטגיה הוא "מיטיגציה" (mitigation) של FVG, כאשר המחיר חוזר לאזור הפער שזוהה וממלא אותו, לפחות חלקית.1 פרמטר 'Threshold %' משמש לסינון גודל ה-FVGs הנבחנים.1
* **LVN (Low Volume Nodes) מפרופיל שוק:**
  + צמתי נפח נמוך (LVNs) מזוהים מניתוח פרופיל שוק על גרפים של 30 דקות.1 צמתים אלו מייצגים רמות מחיר עם נפח מסחר נמוך היסטורית ונחשבים ל"תחנות" משמעותיות או נקודות מפנה פוטנציאליות בהן השוק עשוי להגיב.1
  + היבט קריטי הוא תפיסת "חוזק התחנה". חוזק זה, המוערך על בסיס אינטראקציות מחיר היסטוריות (מבחני תמיכה/התנגדות, נפח שנסחר ב-LVN במהלך בדיקות קודמות), ישפיע ישירות על החלטות גודל הפוזיציה שיתקבלו על ידי סוכני ה-MARL. הדבר מכניס רכיב ניהול סיכונים דינמי ונלמד לתוך האסטרטגיה.1
* **נרות הייקן אשי:**
  + האסטרטגיה משתמשת בנרות הייקן אשי עבור ניתוחי 30 הדקות (MLMI, NW-RQK, וזיהוי משטרי MMD) כדי להפחית רעשי שוק ולספק אינדיקציות מגמה חלקות יותר.1
  + לעומת זאת, ניתוח FVG ב-5 דקות (וככל הנראה גם הנתונים הבסיסיים לפרופיל שוק/LVN, אם כי נקודה זו טעונה אישור) ישתמש בנרות יפניים סטנדרטיים כדי ללכוד תבניות מדויקות של פעולת מחיר.1 ייצוג כפול זה דורש טיפול זהיר בצנרת הנתונים.
* *ספריות מפתח לאינדיקטורים וטיפול בנתונים (רשימה ראשונית):*
  + scikit-learn: למימוש רכיב ה-k-NN של ה-MLMI.1
  + TA-Lib, pandas-ta: לאינדיקטורים טכניים סטנדרטיים כמו RSI ו-WMA, המהווים בסיס ל-MLMI.1
  + smartmoneyconcepts: חבילת פייתון שיכולה לשמש לזיהוי Fair Value Gaps.1
  + marketprofile (ספריית פייתון): מיועדת ספציפית ליצירת פרופילי שוק מנתוני OHLCV, דבר חיוני לזיהוי LVN.1
  + Pandas, Numpy: ספריות יסוד למניפולציה של נתונים, ניתוח סדרות עתיות וחישובים נומריים בפייתון.1

#### **1.1.2. ניתוח רב-ממדי (Multi-Timeframe) וסנכרון אותות**

* האסטרטגיה ממנפת גישה דו-ממדית של מסגרות זמן: גרף ה-30 דקות (MLMI, NW-RQK, LVN, MMD) קובע את ההקשר השוקי הרחב, כיוון המגמה ורמות מבניות משמעותיות.1
* גרף ה-5 דקות (FVG, צורות הרמוניות) משמש לתזמון כניסה מדויק יותר ולניתוח הגיאומטרי החדשני של צירופי אותות.1
* **אתגר הסנכרון:** אתגר טכני מרכזי, המודגש ב-1, הוא ההקרנה או המיפוי המדויק של אותות הנוצרים במסגרת זמן של 30 דקות על גבי גרף ה-5 דקות. הדבר קריטי מכיוון שאותות ה-30 דקות מהווים קודקודים של "המשולשים ההרמוניים" המשורטטים ברזולוציית 5 דקות. בחירת נר ה-5 דקות הספציפי בתוך בר ה-30 דקות המייצג את אות ה-30 דקות (למשל, הנר הראשון, האחרון, או בעל הנפח הגבוה ביותר ב-5 דקות) תשפיע ישירות על הגיאומטריה של המשולשים, וכתוצאה מכך, על המאפיינים הנלמדים על ידי סוכני ה-MARL. עקביות בכלל מיפוי זה היא בעלת חשיבות עליונה. 2 ו-3 נוגעים ביצירת אותות מבוססי-זמן אך אינם פותרים ישירות בעיית מיפוי ספציפית זו, מה שמצביע על כך שמדובר בנקודת לוגיקה מותאמת אישית.

#### **1.1.3. סינרגיות מונעות-שוק וניתוח "צורות הרמוניות" (משולשים)**

* **ארבע הסינרגיות:** אלו אינן אסטרטגיות נפרדות הנבחרות על ידי סוכן, אלא תוצאה של דינמיקת השוק. הן מייצגות את ארבעת הרצפים הזמניים האפשריים והתקפים שבהם שלושת האותות העיקריים (MLMI, NW-RQK, מיטיגציית FVG) יכולים להופיע ולהתיישר ליצירת מערך כניסה שלם.1 לדוגמה, סינרגיה אחת כזו היא אות MLMI ← אות NW-RQK ← מיטיגציית FVG.1 תפקיד סוכן ה-MARL הוא להגיב באופן מיטבי *לאחר* שאחת מהסינרגיות מונעות-השוק הללו התממשה.1
* **צורות הרמוניות (משולשים):**
  + זהו אלמנט חדשני ומרכזי באסטרטגיה. צירוף שלושת האותות העיקריים (MLMI, NW-RQK, מיטיגציית FVG) מתפרש באופן גיאומטרי כיצירת קודקודים של משולש על גרף ה-5 דקות.1
  + **הנדסת מאפיינים גיאומטריים:** מאפיינים גיאומטריים כמותיים מופקים ממשולשים אלו. מאפיינים אלו יכולים לכלול, בין היתר, את אורכי צלעות המשולש, זוויותיו הפנימיות, יחסים בין צלעות או בין גבהים לבסיסים, ושטח המשולש.1
  + **מטרה אסטרטגית:** מאפיינים גיאומטריים אלו נועדו לשמש כמדד ל"איכות" או "תוקף" של מערך הכניסה הפוטנציאלי. הנחת היסוד היא שתצורות גיאומטריות מסוימות של משולשים אלו מתואמות עם הסתברות גבוהה יותר להצלחת העסקה, והערכה זו אמורה להשפיע ישירות על החלטות ניהול הסיכונים, ובפרט על קביעת גודל הפוזיציה.1
  + **נרמול מאפיינים גיאומטריים:** כפי שמודגש ב-1 ונתמך על ידי פרקטיקות ML כלליות 4, מאפיינים גיאומטריים אלו חייבים לעבור נרמול כדי להבטיח שהם ברי-השוואה בתנאי שוק שונים (למשל, רמות מחיר או תנודתיות משתנות). לדוגמה, אורכי צלעות (מימד מחיר) יכולים להיות מנורמלים על ידי ה-Average True Range (ATR) של גרף ה-5 דקות, בעוד שזוויות הן מטבען בלתי תלויות סקאלה.
* *ספריות מפתח לניתוח גיאומטרי:*
  + Numpy: חיוני לחישובים וקטוריים, גיאומטריה אנליטית ופעולות מטריציוניות הנדרשות לחישוב אורכי צלעות, זוויות (למשל, באמצעות מכפלות סקלריות או חוק הקוסינוסים), ושטח.1
  + math (ספריית פייתון סטנדרטית): מספקת פונקציות טריגונומטריות (acos, sqrt) הנחוצות לחישוב זוויות מאורכי צלעות (חוק הקוסינוסים) ונוסחאות גיאומטריות אחרות.1
  + Shapely (אופציונלי): אף שעשוי להיות מורכב מדי עבור משולשים בסיסיים, ספרייה זו יכולה להיות שימושית אם הניתוח הגיאומטרי יורחב בעתיד לצורות מורכבות יותר או ליחסים מרחביים.1

הרובסטיות של האסטרטגיה נשענת על האינטראקציה הסינרגטית והרצף הזמני של סוגי אינדיקטורים מגוונים (מומנטום, רגרסיה, חוסר איזון, פרופיל נפח), ולא על השלמות המבודדת של אינדיקטור בודד. המעבר מ-FVG של דקה 1 ל-FVG של 5 דקות 1 ישנה מטבעו את המרווחים הזמניים בין האותות, ובכך ישנה את הגיאומטריה המתקבלת של "הצורות ההרמוניות". הדבר מרמז שסוכן ה-MARL חייב להיות מסוגל ללמוד ולהסתגל ל"חתימות" גיאומטריות שונות של הזדמנויות הנובעות משינויי תזמון כאלה. שינוי מסגרת הזמן של אות מכונן, כמו FVG, משפיע ישירות על קואורדינטת הזמן של אחד מקודקודי המשולש.1 שינוי בקואורדינטת זמן של קודקוד משנה בהכרח את המאפיינים הגיאומטריים של המשולש (אורכי צלעות, זוויות, שטח).1 מאחר שסוכן ה-MARL משתמש במאפיינים גיאומטריים אלו להערכת איכות העסקה וניהול הסיכונים 1, עליו כעת ללמוד מהתפלגות שונה פוטנציאלית של צורות "חיוביות" לעומת "שליליות". הדבר מדגיש דרישת הסתגלות מובלעת: המערכת חייבת להיות רובסטית לא רק לשינויי שוק חיצוניים אלא גם לשינויי תצורה פנימיים המשפיעים על מאפיינים נגזרים. הסוכן צריך ללמוד שלצורה הנגזרת מ-FVG של 5 דקות עשויות להיות השלכות שונות מאשר לצורה הנגזרת מ-FVG של דקה.

ההחלטה האסטרטגית להשתמש בפרמטרי ברירת מחדל של אינדיקטורים 1 עבור MLMI ו-NW-RQK מעבירה את מלוא נטל הפרשנות האדפטיבית לסוכן ה-MARL, המונחה על ידי זיהוי משטרי MMD. הדבר מפשט את ניהול האינדיקטורים אך מעלה את מורכבות משימת הלמידה עבור הסוכן, ודורש ייצוג מצב עשיר יותר ומדיניות מתוחכמת יותר. פיתוח אסטרטגיה מסורתי כולל לעיתים קרובות כוונון עדין של פרמטרי אינדיקטורים לנכסים ספציפיים או לתנאי שוק.1 על ידי קיבוע הפרמטרים, האותות הגולמיים מאינדיקטורים אלו יפגינו התנהגות עקבית ללא תלות במשטר השוק השורר. סוכן ה-MARL מקבל את משטר השוק שזוהה על ידי MMD כחלק מקלט המצב שלו.1 כתוצאה מכך, הסוכן חייב ללמוד שאות ספציפי (למשל, חציית MLMI) מאינדיקטור בעל פרמטרים קבועים עשוי לסמן עסקה בעלת הסתברות גבוהה ב"משטר A" אך עסקה בעלת הסתברות נמוכה ב"משטר B". הדבר מחייב מדיניות מורכבת יותר מצד סוכן ה-MARL, שכן עליו כעת לבצע את הפרשנות ההקשרית שאחרת הייתה עשויה להיות מטופלת חלקית על ידי כוונון אינדיקטורים דינמי. משמעות הדבר היא שייצוג המצב של הסוכן חייב להיות עשיר מספיק כדי לאפשר הבחנה זו.

השימוש הדיפרנציאלי בנרות הייקן אשי עבור אינדיקטורים של 30 דקות (MLMI, NW-RQK, MMD) 1 ובנרות יפניים סטנדרטיים עבור ניתוח FVG ב-5 דקות 1 מציג הטרוגניות בייצוג הנתונים. בעוד שנרות הייקן אשי מחליקים מגמות, הם עלולים להסתיר או לעכב את תבניות פעולת המחיר המדויקות שעליהן מסתמך זיהוי FVG, או לשנות את התפלגות הנפח הנתפסת עבור ניתוח LVN אם גם LVNs נגזרים מהייקן אשי. היישור והפרשנות של אותות מסוגי נרות שונים אלו דורשים טיפול קפדני בהנדסת המאפיינים ובייצוג המצב כדי להבטיח עקביות עבור סוכן ה-MARL. נרות הייקן אשי (HA) מחושבים על ידי מיצוע נתוני בר קודמים, מה שמביא למגמות חלקות יותר אך לערכי OHLC שונים בהשוואה לנרות יפניים סטנדרטיים.1 MLMI, NW-RQK, ו-MMD מיועדים להשתמש בנרות HA של 30 דקות.1 ניתוח FVG מיועד לנרות סטנדרטיים של 5 דקות.1 ניתוח LVN משתמש בדרך כלל בנתוני OHLCV סטנדרטיים ליצירת פרופיל נפח מדויק. זיהוי FVG סטנדרטי תלוי ביחסים ספציפיים בין ערכי השיא/שפל הגולמיים של שלושה נרות סטנדרטיים עוקבים.1 יישום לוגיקה זו ישירות על נרות HA יזהה אזורי FVG שונים (או לא יזהה כלל) בשל אפקט המיצוע על ערכי השיא והשפל. פרופיל שוק (עבור LVN) צובר נפח ברמות מחיר בדידות הנגזרות מנרות סטנדרטיים. רמות המחיר של נרות HA, בהיותן ממוצעות, יעוותו פרופיל זה. אם אותות מאינדיקטורים מבוססי-HA (MLMI, NW-RQK) משולבים עם אותות מ-FVG/LVN מבוססי-נרות-סטנדרטיים ליצירת "משולשים הרמוניים", קואורדינטות המחיר של הקודקודים יגיעו מייצוגי נתונים שונים. הדבר עלול להכניס חוסר עקביות עדין או לדרוש נרמול/פרשנות מתוחכמים על ידי סוכן ה-MARL. לדוגמה, נקודת מיטיגציה של FVG על גרף נרות סטנדרטיים עשויה להתאים לרמת מחיר פחות מובחנת על גרף HA לאותה תקופה, מה שעלול להשפיע על הגיאומטריה הנתפסת או על משמעותה. המערכת חייבת ליישב ייצוגים אלו.

טבלה 1.1 (חלק I, סעיף 1.1): סקירת רכיבי ליבה אסטרטגיים

(מבוסס על 1)

| **שם הרכיב** | **מסגרת זמן (מקור נתונים)** | **סוג נרות קלט** | **סיכום לוגיקה/חישוב ליבה** | **אות/מאפיין לאסטרטגיה** | **תפקיד באסטרטגיה/צורה הרמונית** |
| --- | --- | --- | --- | --- | --- |
| MLMI (Machine Learning Momentum Index) | 30 דקות | הייקן אשי | שילוב RSI ו-k-NN לתחזית מומנטום פן וון אדפטיבית | חציית קו תחזית MLMI עם ה-WMA שלו | אות ראשוני/מאשר בארבע הסינרגיות; קודקוד במשולש ההרמוני |
| NW-RQK (Nadaraya-Watson Rational Quadratic Kernel Regression) | 30 דקות | הייקן אשי | רגרסיית ליבה לא-פרמטרית ליצירת עקומת התאמה למחיר (non-repainting) | שינוי כיוון עקומת הרגרסיה (שורי/דובי) | אות ראשוני/מאשר בארבע הסינרגיות; קודקוד במשולש ההרמוני |
| FVG (Fair Value Gap) - LuxAlgo | 5 דקות | נרות סטנדרטיים | זיהוי פערי מחיר המייצגים חוסר איזון | "מיטיגציה" של FVG (חזרת המחיר למילוי הפער) | אות מאשר בארבע הסינרגיות; קודקוד במשולש ההרמוני |
| LVN (Low Volume Nodes) | פרופיל שוק מנתוני OHLCV של 30 דקות | נרות סטנדרטיים | זיהוי אזורי מחיר עם נפח מסחר נמוך היסטורית | "עוצמת תחנה" של LVN (מבוסס בדיקות קודמות, נפח) | משפיע על הערכת איכות כניסה ועל גודל פוזיציה (ניהול סיכונים) |
| משולש הרמוני | 5 דקות (נוצר משילוב אותות MLMI, NW-RQK, FVG) | נרות סטנדרטיים | ייצוג גיאומטרי של צירוף שלושת האותות | מאפיינים גיאומטריים (זוויות, יחסי צלעות, שטח וכו') | משמש להערכת "איכות" הכניסה ולהשפיע על גודל הפוזיציה |
| MMD Market Regime | 30 דקות | הייקן אשי | זיהוי שינויים התפלגותיים בנתוני שוק באמצעות Higher-Rank MMD וחתימות נתיבים | תווית מצב שוק (קטגוריאלית) או ציון MMD (רציף) | מספק הקשר שוק גלובלי לסוכני ה-MARL, מאפשר התאמה אדפטיבית של פרשנות האותות וההחלטות |

טבלה זו מספקת נקודת ייחוס ברורה ומרוכזת לכל אבני הבניין של האסטרטגיה, מסגרות הזמן שלהן, ואופן תרומתן של אותות או מאפיינים לתהליך קבלת ההחלטות. היא מהווה בסיס להבנת תכנון צנרת הנתונים ומטריצת הקלט בהמשך.

### **סעיף 1.2: צנרת נתונים ובניית מטריצת קלט**

חלק זה מפרט את המסלול הקריטי מהמרת נתוני שוק גולמיים ועד ליצירת מטריצת קלט מובנית ועשירת-מאפיינים, המוכנה להזנה לשכבת ההטמעה של הרשת הנוירונית של מערכת ה-MARL. תהליך זה כולל רכישת נתונים, עיבוד מקדים, חישוב מאפיינים, והרכבה מחושבת של סוגי נתונים מגוונים.

#### **1.2.1. רכישת נתוני OHLCV ועיבודם המקדים עבור שוקי החוזים העתידיים**

* **מקורות נתונים:** הנתונים העיקריים יהיו נתוני OHLCV עבור חוזים עתידיים. המשתמש ציין את Rithmic כמקור אפשרי 1, בעוד ש-Databento ו-dxFeed הן ספקיות API מומלצות הידועות באיכות נתוני החוזים העתידיים שלהן ובממשקי API ידידותיים למפתחים.1 המערכת נמנעת במפורש משימוש בנתוני Level 2 כדי להפחית רעש ומורכבות.1
* **היקף נתונים היסטוריים:** נדרש מינימום של חמש שנות נתונים היסטוריים לצורך בדיקות לאחור חזקות ואימון ראשוני של המודלים (משתמע מרצון לבדיקות מקיפות, אף ש-1 מציין שנתיים).
* **מסגרות זמן וגרנולריות:** הנתונים יעובדו באינטרוולים של 30 דקות ו-5 דקות.1
* **שלבי עיבוד מקדים:**
  + **המרת נרות הייקן אשי:** נתוני OHLCV של 30 דקות (עבור MLMI, NW-RQK, MMD) יומרו לנרות הייקן אשי כדי להחליק מגמות ולהפחית רעש.1 נתוני ה-5 דקות לניתוח FVG יישארו כנרות יפניים סטנדרטיים.1 ההחלטה לגבי סוג הנרות עבור נתוני ה-5 דקות המשמשים לשרטוט המשולשים ההרמוניים צריכה להיות עקבית עם אות ה-FVG, ולכן סביר שתהיה מבוססת על נרות סטנדרטיים, או על נרות הייקן אשי אם ניתן להפיק אות FVG אמין מהם. נקודה זו דורשת שיקול דעת זהיר להבטחת עקביות.
  + **ניקוי נתונים:** נהלים סטנדרטיים לטיפול בנקודות נתונים חסרות (למשל, אינטרפולציה, מילוי קדימה, אם כי פחות קריטי עם אופי המיצוע של הייקן אשי) וזיהוי/טיפול בחריגות (שוב, החלקת הייקן אשי עשויה למתן חלק מהחריגות) הם הכרחיים.1
  + **ניהול חותמות זמן:** הבטחת דיוק חותמות זמן וסנכרון בין זרמי הנתונים של 30 ו-5 דקות היא בעלת חשיבות עליונה ליישור אותות נכון ולהנדסת מאפיינים.1
  + **חוזים מתגלגלים (Continuous Contracts):** מימוש לוגיקה לטיפול בגלגול חוזים עתידיים חיוני ליצירת סדרת נתונים רציפה וארוכת טווח לבדיקות לאחור ואימון.1 Databento מציעה "Smart Symbology" למטרה זו.1

#### **1.2.2. תכנון מטריצת הקלט NxF להזנת הסוכנים**

* **מבנה המטריצה:** הקלט לשכבת ההטמעה של סוכן ה-MARL יהיה מטריצה במבנה של N×F, כאשר N מייצג את אורך הרצף (מספר צעדי זמן היסטוריים) ו-F הוא מספר המאפיינים בכל צעד זמן.1
* **גרנולריות צעד הזמן:** מאחר שהמשולשים ההרמוניים מנותחים על גרף של 5 דקות, הגיוני שצעד הזמן הבסיסי של המטריצה יהיה 5 דקות.1
* **אורך הרצף (N):** זהו היפר-פרמטר קריטי הקובע את חלון ההסתכלות לאחור. בחירתו כרוכה באיזון בין אספקת הקשר היסטורי מספק למודל (במיוחד עבור מאפייני 30 דקות הנעים לאט יותר) לבין ניהול מורכבות חישובית והסיכון להתאמת יתר. 1 דן בכך, וספרות (למשל, הפניות 6, 7 ב-1) מראה ערכי N משתנים בהתאם לתדירות הנתונים. עבור נתוני 5 דקות, N יכול לנוע בין מספר שעות (למשל, N=24 עבור שעתיים) ועד לסשן מסחר מלא או יותר.
* **הרכב וקטור המאפיינים (F מאפיינים לכל צעד זמן של 5 דקות):** בהתבסס על 1 ו-1:
  + **נתוני נרות 5 דקות:** OHLCV (סטנדרטיים או הייקן אשי, בהתאם להחלטה על שרטוט FVG/משולשים).
  + **נתוני נרות הייקן אשי של 30 דקות (הקשר):** ה-OHLCV של בר ההייקן אשי של 30 דקות שאליו שייך בר ה-5 דקות הנוכחי. ערך זה נישא קדימה למשך שישה ברי 5 דקות.
  + **ערכי אינדיקטורים (מיושרים לצעדי 5 דקות):**
    - MLMI (מ-30 דקות HA): ערך נוכחי של קו MLMI, WMA של MLMI (נישא קדימה).
    - NW-RQK (מ-30 דקות HA): ערך נוכחי של עקומת הרגרסיה, שיפוע/כיוון (נישא קדימה).
    - FVG (מ-5 דקות, פוטנציאלית מבוסס HA): דגלים בינאריים לקיום FVG שורי/דובי, מרחק לגבולות FVG, סטטוס מיטיגציה.
  + **מאפייני LVN (מיושרים לצעדי 5 דקות):** קרבה (למשל, מרחק בטיקים או מכפלות ATR) ל-LVN המשמעותי הקרוב ביותר שזוהה בפרופיל של 30 דקות. ציון "חוזק תחנה" מכומת עבור LVN(s) רלוונטי(ים) אלו (נישא קדימה).
  + **מאפייני משולש הרמוני (מבוססי-אירוע, מיושרים לצעדי 5 דקות):** תכונות גיאומטריות (זוויות, אורכי צלעות מנורמלים, שטח) של המשולש ההרמוני האחרון שהושלם. זמן שחלף (בברי 5 דקות) מאז היווצרות המשולש האחרון. דגל בינארי: 1 אם משולש חדש הושלם בבר 5 הדקות הנוכחי, 0 אחרת.
  + **פלט מצב שוק MMD (מיושר לצעדי 5 דקות):** תווית מצב נוכחי (למשל, מקודדת one-hot) או ציון MMD רציף (נישא קדימה).
* **נרמול וסקיילינג:** חיוני ליציבות וביצועי הרשת הנוירונית. כל אחד מ-F המאפיינים צריך לעבור סקיילינג מתאים (למשל, סטנדרטיזציית Z-score עבור תשואות או אינדיקטורים בלתי חסומים, סקיילינג Min-Max עבור אינדיקטורים חסומים כמו RSI או זוויות [0-180/0−π]) על פני מערך נתוני האימון.1
* **מידע זמני היררכי בווקטור המאפיינים:** וקטור המאפיינים משלב מידע מרזולוציות זמן שונות. מאפיינים מחישובי 30 דקות (MLMI, NW-RQK, LVN, MMD) יהיו קבועים למשך שש שורות עוקבות של 5 דקות במטריצה, ואז יתעדכנו. אופי קבוע-למקוטעין זה הוא בעצמו דפוס שהרשת הנוירונית יכולה ללמוד.1 שכבת ההטמעה של הרשת הנוירונית (למשל, LSTM או Transformer) תצטרך ללמוד שלמאפיינים אלו תדירות עדכון שונה והם מייצגים הקשר הנע לאט יותר בהשוואה לנתוני FVG או הייקן אשי של 5 דקות. תזמון היררכי זה הוא היבט חיוני של ניתוח רב-מסגרות-הזמן.

#### **1.2.3. טיפול במאפיינים מבוססי-אירועים וספורדיים בתוך רצפים באורך קבוע**

* **האתגר:** משולשים הרמוניים הם מונעי-אירועים; הם אינם נוצרים בכל אינטרוול של 5 דקות. גם מיטיגציות FVG הן אירועים. מטריצת ה-N×F, לעומת זאת, דורשת מערך עקבי של F מאפיינים בכל צעד זמן N.
* **אסטרטגיית ייצוג למשולשים הרמוניים:**
  + **דגל אירוע נוכחי:** מאפיין בינארי המציין האם משולש הרמוני חדש הושלם בצעד הזמן הנוכחי של 5 דקות.
  + **מאפייני האירוע האחרון:** אחסון התכונות הגיאומטריות (זוויות, אורכי צלעות מנורמלים וכו') של המשולש האחרון שהושלם. ערכים אלו נישאים קדימה עד להיווצרות משולש חדש.
  + **זמן מאז האירוע האחרון:** מאפיין מספרי הסופר את מספר ברי ה-5 דקות מאז הושלם המשולש ההרמוני האחרון. הדבר מסייע למודל להעריך את "טריות" או רלוונטיות המאפיינים הגיאומטריים הנישאים קדימה.1
  + ספרות כגון 1 (הפניה 11) דנה בטיפול באירועים נדירים ובבחירת מודלים בהתבסס על תצפיות סדרות זמן, ומדגישה את הצורך בייצוג חזק.
* **אסטרטגיית ייצוג לאותות FVG:**
  + **סטטוס אזור FVG:** מאפיינים המציינים האם קיים אזור FVG שורי/דובי, המרחק לגבולות הפער, סטטוס מיטיגציה (למשל, לא התרחשה, התרחשה חלקית, התרחשה במלואה). מאפיינים אלו יכולים להתעדכן בכל צעד זמן של 5 דקות בהתבסס על תנועת המחיר ביחס לפערים מזוהים.

בחירת הייצוג לאירועים ספורדיים יכולה להשפיע משמעותית על הלמידה. נשיאת נתוני אירוע ישנים קדימה עשויה לספק הקשר אך גם להכניס מידע לא רלוונטי אם לא מנוהלת עם מאפיין טריות. ערכי מילוי-מקום דורשים מהרשת הנוירונית ללמוד להתעלם מהם או לפרשם נכונה.1 אם משתמשים במאפייני המשולש הידוע האחרון, הסוכן מקבל קלט גיאומטרי רציף, אך הוא עשוי להיות מאירוע ישן ולא רלוונטי. הוספת "זמן מאז האירוע" מסייעת לסוכן להפחית את משקלם של נתונים גיאומטריים ישנים. לחלופין, ריפוד באפסים או שימוש בדגלים מיוחדים עבור "אין אירוע נוכחי" הופכים את המאפיינים הגיאומטריים לספורדיים באמת בווקטור הקלט, והרשת הנוירונית חייבת ללמוד להתמודד עם כך. החלטה זו משפיעה על מורכבות מה ששכבת ההטמעה וסוכן ה-RL צריכים ללמוד.

טבלה 1.2: סכמת מטריצת קלט מאוחדת (מבנה N×F רעיוני עבור צעד זמן בודד)

(מבוסס על 1)

| **קטגוריית מאפיין** | **שם מאפיין ספציפי** | **סוג נתונים** | **שיטת נרמול/סקיילינג מוצעת** |
| --- | --- | --- | --- |
| **נתוני נרות (5 דקות)** | HA\_Open\_5m, HA\_High\_5m, HA\_Low\_5m, HA\_Close\_5m (או סטנדרטי, תלוי בהחלטת FVG/משולש) | רציף | אחוז שינוי / לוג-תשואה, ואז סטנדרטיזציה |
|  | Volume\_5m | רציף | סטנדרטיזציה או Min-Max (אם חסום) |
| **נתוני נרות (הקשר 30 דקות)** | HA\_Open\_30m, HA\_High\_30m, HA\_Low\_30m, HA\_Close\_30m | רציף | כנ"ל (ערך נישא קדימה מבר 30 דקות אחרון) |
|  | HA\_Volume\_30m | רציף | כנ"ל (ערך נישא קדימה) |
| **אינדיקטורים (מיושרים ל-5 דקות)** | MLMI\_Value, MLMI\_WMA\_Value (מ-30 דקות HA) | רציף | Min-Max (אם האינדיקטור חסום) או סטנדרטיזציה (ערך נישא קדימה) |
|  | NWRQK\_Value, NWRQK\_Slope (מ-30 דקות HA) | רציף | סטנדרטיזציה (ערך נישא קדימה) |
|  | FVG\_Bull\_Active, FVG\_Bear\_Active (מ-5 דקות) | בינארי | - |
|  | FVG\_Distance\_Upper, FVG\_Distance\_Lower (מ-5 דקות) | רציף | נרמול לפי ATR או טווח מחירים |
|  | FVG\_Mitigation\_Status (מ-5 דקות) | קטגוריאלי (למשל, 0=אין, 1=חלקי, 2=מלא) | One-hot encoding |
| **מאפייני LVN (מיושרים ל-5 דקות)** | LVN\_Nearest\_Distance, LVN\_Nearest\_Strength\_Score (מפרופיל 30 דקות) | רציף | נרמול לפי ATR; Min-Max לציון (ערך נישא קדימה) |
| **גיאומטריית משולש הרמוני** | Triangle\_Angle\_A, Triangle\_Angle\_B, Triangle\_Angle\_C | רציף (0-180) | Min-Max לטווח $\[0, \pi]$ או $\[-1, 1]$ (לאחר המרה לקוסינוס/סינוס) |
|  | Triangle\_Side\_Ratio\_12, Triangle\_Side\_Ratio\_23 (אורכי צלעות מנורמלים) | רציף | סטנדרטיזציה (אם הטווח אינו מוגדר היטב לאחר נרמול ראשוני של צלעות) |
|  | Triangle\_Area\_Normalized | רציף | Min-Max לאחר נרמול ראשוני (למשל, חלקי ATR בריבוע) |
|  | Triangle\_Time\_Since\_Formation | שלם (מספר ברים) | Min-Max על פני חלון היסטורי סביר או לוג-טרנספורמציה |
|  | Triangle\_Newly\_Formed\_Flag | בינארי | - |
| **מצב שוק MMD (מיושר ל-5 דקות)** | MMD\_Regime\_Label (אם קטגוריאלי, מ-30 דקות HA) | קטגוריאלי | One-hot encoding (ערך נישא קדימה) |
|  | MMD\_Score (אם רציף, מ-30 דקות HA) | רציף | סטנדרטיזציה (ערך נישא קדימה) |

טבלה זו עונה ישירות על שאלת הליבה של המשתמש לגבי סכמת המטריצה. היא מספקת תוכנית קונקרטית להנדסת מאפיינים ובניית צנרת נתונים, ומבטיחה שכל הרכיבים האסטרטגיים שצוינו יילקחו בחשבון בקלט לרשת הנוירונים. היא מבהירה את הממדיות (F) של וקטור המאפיינים בכל צעד זמן.

## **חלק II: זיהוי מצבי שוק מתקדם והטמעת קלט סוכן**

חלק זה מתמקד בשני רכיבים קריטיים של מערכת המסחר: מנגנון זיהוי מצבי השוק, המספק הקשר מאקרו-כלכלי חיוני, והאופן שבו מטריצת הקלט המורכבת N×F מעובדת לייצוג וקטורי דחוס ואינפורמטיבי עבור סוכני ה-MARL.

### **סעיף 2.1: זיהוי מצבי שוק באמצעות Higher-Rank MMD**

היכולת לזהות ולהגיב למצבי שוק משתנים היא אבן יסוד באסטרטגיות מסחר אלגוריתמיות מודרניות. שווקים פיננסיים אינם סטציונריים, והתנהגותם יכולה לעבור בין משטרים שונים כגון מגמה ברורה, תנודתיות גבוהה ללא כיוון מוגדר, או דשדוש צר. אסטרטגיית מסחר שתתפקד היטב במשטר אחד עלולה להיכשל במשטר אחר. לכן, זיהוי משטרים דינמי מאפשר למערכת להתאים את התנהגותה, כגון תעדוף רצפי אותות מסוימים, התאמת פרמטרי כניסה/יציאה, או שינוי אסטרטגיות ניהול סיכונים.1

#### **2.1.1. עקרונות תיאורטיים: חתימות נתיבים, שיטות ליבה ו-MMD מסדר גבוה**

* **רציונל לבחירת MMD:** המערכת תשתמש במבחן הלא-פרמטרי Maximum Mean Discrepancy (MMD) לצורך השוואת התפלגויות של נתוני שוק.1 המשתמש הביע העדפה לגרסה מתקדמת, Higher-Rank MMD, בשל יכולתה ללכוד תלות בנתיב (path-dependency) ודינמיקות לא-מרקוביות באמצעות שימוש בחתימות נתיבים מחוספסים (rough path signatures).1
* **חתימות נתיבים (Path Signatures):** אלו הם ייצוגים מתמטיים של סדרות זמן הלוכדים את הגיאומטריה שלהן ואת סדר התנועות באופן שאינו תלוי בפרמטריזציה מחדש של הזמן.1 החתימות יופקו ממאפייני נרות הייקן אשי (למשל, לוג-תשואות, תנודתיות) במסגרת זמן של 30 דקות.1
* **שיטות ליבה (Kernel Methods):** ליבת חתימה (signature kernel) תשמש בשילוב עם MMD להשוואת התפלגויות של נתיבים במרחב הילברט עם ליבה משחזרת (Reproducing Kernel Hilbert Space - RKHS).1
* **Higher-Rank MMD:** הרחבה זו של MMD משלבת מידע פילטרציה (התפתחות המידע לאורך זמן), מה שהופך אותה לרגישה יותר לשינויים עדינים בדינמיקת השוק.1
* **זיהוי מקוון (Online Detection):** השוואת חלון נתונים עדכני לחלון ייחוס היסטורי לצורך זיהוי שינויים בזמן אמת.1

הבחירה ב-"Higher-Rank MMD" היא מתוחכמת ועתירת חישוב. יתרונה המרכזי הוא הרגישות למאפיינים תלויי-נתיב ולא-מרקוביים, המאפיינים שווקים פיננסיים. האתגר המעשי יהיה כיול הפרמטרים שלה לאיזון אופטימלי בין תגובתיות לתוצאות חיוביות שגויות (false positives), במיוחד בהינתן תשומות הייקן אשי המוחלקות.1 החלקת הייקן אשי עשויה להקשות על זיהוי שינויים התפלגותיים עדינים, מה שעשוי לדרוש פרמטרי MMD רגישים יותר. האיזון הוא בין לכידת שינויי מצב אמיתיים במהירות לבין הימנעות מהפעלה על ידי רעש או תנודות קלות המוגברות על ידי הגדרות רגישות.

פלט ה-MMD (ציון מרחק או תווית מצב) הוא סיכום בעל ממדיות נמוכה יחסית של השוואה התפלגותית מורכבת. סוכן ה-RL לא יראה את חישוב ה-MMD הגולמי אלא את תוצאתו, מה שמפשט את מרחב המצבים של הסוכן אך גם מפשיט את הניואנסים של השינוי ההתפלגותי. ערך/תווית יחיד זה מוזן למטריצת הקלט הגדולה הרבה יותר עבור סוכן ה-RL. הסוכן לומד לקשר פלט MMD זה לפעולות אופטימליות. משמעות הדבר היא שהסוכן אינו לומד *מדוע* מצב השוק השתנה, אלא רק *שהוא* השתנה וכיצד להגיב. זוהי הפשטה המקלה על משימת הסוכן אך מסתמכת במידה רבה על יכולתו של מנוע ה-MMD לזהות נכונה שינויים משמעותיים.

#### **2.1.2. פרטי מימוש מנוע זיהוי המשטרים**

* **ספריות וקוד מקור:** המימוש המעשי של מנוע זיהוי המשטרים יתבסס על מאגר הקוד הפתוח issaz/signature-regime-detection 1 וחבילת הפייתון higherOrderKME.1 מאגרים אלו מספקים קוד פייתון ליישום sig-MMD לזיהוי משטרים באופן מקוון, וכוללים מחברות Jupyter עם דוגמאות רלוונטיות (כגון 4-online-regime-detection.ipynb ו-4-higher-rank-mmd.ipynb).
* **פרמטרים מרכזיים לתצורה וכיול:**
  + **סוג הגרעין (Kernel Type):** למשל, Gaussian Kernel על מרחב החתימות.1
  + **פרמטרי הגרעין (למשל, רוחב פס σ):** ייקבע באמצעות כיול (למשל, cross-validation, היוריסטיקת חציון).1
  + **רמת קיטוע החתימה (M):** למשל, 2 או 3 (ערכים נמוכים יותר לחישוב מהיר יותר).1
  + **גודל חלון היסטורי (רפרנס):** למשל, 60-120 נרות של 30 דקות (30-60 ימי מסחר).1
  + **גודל חלון נוכחי (לבדיקה):** למשל, 20-30 נרות של 30 דקות (10-15 ימי מסחר).1
  + **סף סטטיסטי לזיהוי שינוי (p-value או ערך MMD):** ייקבע אמפירית או באמצעות מבחני מובהקות.1
  + **תדירות עדכון:** קביעת התדירות שבה יבוצע חישוב ה-MMD מחדש (למשל, כל נר חדש של 30 דקות).1
  + פרמטרים אלו מפורטים ב-1 (טבלה 2) ודורשים כיול קפדני. העלות החישובית של MMD מסדר גבוה, במיוחד בשילוב עם חתימות נתיב, יכולה להיות משמעותית.1 יש למצוא איזון בין יכולת לכידת מידע פילטרציה מורכב לבין ההשהיה (latency) בזיהוי המשטר, שהיא קריטית עבור מערכת הפועלת בזמן אמת. ייתכן שיידרשו אופטימיזציות של המימוש או בחירה מושכלת של סדר ה-MMD ורמת קיטוע החתימה כדי להבטיח תגובה מהירה מספיק.

#### **2.1.3. שילוב פלט זיהוי המשטרים במצב הסוכן**

הפלט ממנוע ה-MMD, שיכול להיות תווית קטגוריאלית המציינת את משטר השוק הנוכחי (למשל, "משטר 1", "משטר 2", או תיאור איכותי יותר אם ניתן לגזור אותו) או ערך MMD רציף (המודד את המרחק בין ההתפלגויות), ישולב כמאפיין נוסף בקלט של כל סוכן MARL.1 שילוב זה יתבצע לפני שלב החלוקה הדינמית של הנתונים לתצוגות (Mi​) של הסוכנים. בדרך זו, כל סוכן, על אף שהוא עשוי לקבל רק מקטע חלקי של נתוני השוק, יהיה מודע להקשר המאקרו-כלכלי הרחב יותר שמספק מנוע זיהוי המשטרים.1

ניתן להעשיר את המידע המסופק לסוכן על ידי שימוש במאפיינים היסטוריים של המשטר שזוהה. לדוגמה, אם מנוע ה-MMD מזהה "משטר X", ניתן לשלוף מ"טבלת מאפייני משטרים" (שתיבנה מניתוח היסטורי) פרמטרים כגון תנודתיות ממוצעת, עוצמת מגמה טיפוסית, או רצפי אותות שהיו רווחיים במיוחד במשטר X בעבר. מידע זה יכול להוות קלט נוסף לסוכן ה-RL, ולאפשר לו להתאים את מדיניותו בצורה מדויקת יותר למשטר הנוכחי.1

אם פלט ה-MMD הוא ציון רציף, הוא עשוי לספק אינדיקציה מורכבת יותר לגבי עוצמת המעבר בין מצבים או רמת הביטחון בזיהוי, ששכבת ההטמעה של הרשת הנוירונית תוכל פוטנציאלית ללמוד לפרש. תווית קטגוריאלית פשוטה יותר אך מאבדת ניואנס זה.1 הבחירה משפיעה על מה ששכבת ההטמעה וסוכן ה-RL צריכים ללמוד ממאפיין ספציפי זה.

טבלה 2.1: תצורת מנוע זיהוי מצבי שוק (MMD)

(מבוסס על 1)

| **פרמטר** | **ערך נבחר / אסטרטגיה** | **רציונל / השפעה על הזיהוי** |
| --- | --- | --- |
| מאפייני קלט לחתימת נתיב | לוג-תשואות ותנודתיות מחושבת (מנרות הייקן אשי של 30 דקות) | לכידת דינמיקת מחיר ותנודתיות כבסיס לזיהוי שינויים התפלגותיים. |
| רמת קיטוע חתימה (M) | למשל, 2 או 3 | קובעת את עומק לכידת התלות הסדרתית; ערכים גבוהים יותר לוכדים יותר מידע אך מגדילים מורכבות חישובית. |
| סוג ליבה (Kernel) עבור MMD | למשל, ליבה גאוסיאנית (RBF) על מרחב החתימות | בחירה נפוצה וגמישה; דורשת כיול פרמטרי ליבה (למשל, רוחב פס σ). |
| גודל חלון ייחוס היסטורי | למשל, 60-120 נרות של 30 דקות (30-60 ימי מסחר) | צריך להיות ארוך מספיק לייצג מצב יציב, אך לא ארוך מדי כדי לא לכלול מצבים ישנים מדי. |
| גודל חלון נוכחי (לבדיקה) | למשל, 20-30 נרות של 30 דקות (10-15 ימי מסחר) | צריך להיות רגיש מספיק לשינויים, אך לא קצר מדי כדי למנוע רעש. |
| סף MMD סטטיסטי לזיהוי שינוי | ייקבע אמפירית או באמצעות מבחני מובהקות (למשל, p-value) | קובע את רמת הביטחון הנדרשת להכרזה על שינוי מצב. סף נמוך מדי -> אזעקות שווא; סף גבוה מדי -> פספוס שינויים. |
| תדירות עדכון חישוב MMD | כל נר חדש של 30 דקות | מאזן בין עדכניות המידע לעלות החישובית. |

טבלה זו חיונית לתיעוד החלטות התכנון של רכיב מורכב וקריטי זה, ומסייעת בשחזור, ניפוי שגיאות וכיול עתידי של מנוע זיהוי מצבי השוק. היא מפרטת את ההחלטות התכנוניות הקריטיות למנוע ה-MMD, המשפיע על יכולת ההסתגלות של המערכת כולה לתנאי שוק משתנים.

### **סעיף 2.2: הפחתת ממדיות וייצוג וקטורי באמצעות רשתות נוירונים**

חלק זה מתעמק באופן שבו מטריצת הקלט הגבוה-ממדית וההטרוגנית N×F עוברת טרנספורמציה לייצוג וקטורי דחוס, המותאם לקלט של סוכני למידת החיזוק.

#### **2.2.1. רציונל ומטרות להטמעת מטריצת הקלט**

* **הבעיה:** מטריצת הקלט N×F, כפי שתוארה, עלולה להיות בעלת ממדיות גבוהה ולהכיל סוגי נתונים הטרוגניים (מאפיינים רציפים, קטגוריאליים, מבוססי-אירועים עם סקלות והתפלגויות שונות).1
* **מטרה 1: הפחתת ממדיות:** יצירת ייצוג וקטורי רציף, דחוס ובעל ממדיות נמוכה יותר.1
* **מטרה 2: הפחתת הטרוגניות:** מיפוי מאפיינים מגוונים למרחב משותף ואחיד יותר, שבו ניתן ללמוד יחסים ביניהם בקלות רבה יותר.1
* **מטרה 3: למידת מאפיינים (Feature Learning):** למידה אוטומטית של צירופים רלוונטיים ואינטראקציות לא-ליניאריות בין מאפייני הקלט הגולמיים.1
* **תועלת ללמידת חיזוק (RL):** אספקת ייצוג מצב יציב יותר, אינפורמטיבי ובעל ממדיות נמוכה יותר עבור סוכני ה-RL, מה שעשוי להאיץ את הלמידה ולשפר את איכות המדיניות הנלמדת.1

רשת ההטמעה פועלת כממשק חיוני בין מרחב המאפיינים המהונדס והמורכב של מטריצת הקלט לבין רשתות המדיניות/ערך של סוכן ה-RL. יעילותה בדחיסת מידע והדגשת דפוסים רלוונטיים היא בעלת חשיבות עליונה. סוכני RL עלולים להתקשות עם מרחבי מצב גבוהי-ממדיות והטרוגניים מאוד ("קללת הממדיות"). רשת נוירונים להטמעה יכולה להקרין זאת לווקטור קטן ודחוס. וקטור זה אמור ללכוד את המידע הבולט ביותר למשימת המסחר, ובכך להפוך את בעיית הלמידה של סוכן ה-RL לקלה יותר לניהול.

#### **2.2.2. ניתוח השוואתי של ארכיטקטורות רשת נוירונים מועמדות**

בחירת הארכיטקטורה המתאימה לשכבת ההטמעה תלויה במאפייני הנתונים ובמטרות הלמידה.

* **רשתות קונבולוציה חד-ממדיות (1D-CNNs):**
  + **יתרונות:** יעילות בזיהוי דפוסים מקומיים ומוטיבים ברצפים.1 יכולות ללמוד מאפיינים היררכיים. יעילות חישובית יחסית.
  + **חסרונות:** שדה קליטה (receptive field) מוגבל, עשויות להתקשות בלכידת תלויות ארוכות-טווח אלא אם משתמשים בקונבולוציות מורחבות (dilated convolutions).1
  + **יישום:** יכולות לסרוק את מטריצת ה-N×F לאורך ציר הזמן (N) כדי לחלץ דפוסים זמניים מ-F המאפיינים, או על פני המאפיינים אם קיים סידור "מרחבי" משמעותי. 1 (הפניה 10) מציין במפורש שימוש ב-1D CNNs להטמעת "טוקנים" (מאפיינים) ברשתות טרנספורמר עבור סדרות זמן פיננסיות, תוך למידת מידע זמני לכל נייר ערך/מאפיין. דוגמאות למימוש 1D-CNN ב-PyTorch ניתן למצוא ב-.8
* **רשתות רקורנטיות (LSTMs/GRUs):**
  + **יתרונות:** מתוכננות במיוחד עבור נתונים סדרתיים, יכולות ללכוד תלויות זמניות ארוכות-טווח והקשר.1 יכולות להתמודד עם רצפים באורך משתנה (אם כי כאן N קבוע).
  + **חסרונות:** עלולות להיות איטיות לאימון בשל עיבוד סדרתי, בעיות פוטנציאליות של התפוגגות/התפוצצות גרדיאנטים (אם כי מופחתות על ידי ארכיטקטורת LSTM/GRU).1 העיבוד הסדרתי מגביל מקבול.1
  + **יישום:** עיבוד רצף N וקטורי המאפיינים כדי לייצר מצב חבוי סופי או רצף של מצבים חבויים כהטמעה. 11 מתאר שימוש ב-LSTM להפקת מאפייני מצב שוק עבור סוכן DDPG.
* **רשתות טרנספורמר (Transformers):**
  + **יתרונות:** משתמשות במנגנוני קשב (attention) ללכידת תלויות גלובליות ברצפים, ניתנות למקבול גבוה, ונחשבות למצב-האמנות (state-of-the-art) במשימות רצף רבות.1 גרסאות כמו ProbSparse Attention יכולות להתמודד ביעילות עם רצפים ארוכים.1
  + **חסרונות:** דורשות כמויות גדולות של נתונים לאימון, עלולות להיות יקרות חישובית עבור רצפים ארוכים מאוד (אם כי קיימות גרסאות יעילות), ועשויות לדרוש כיול קפדני.1
  + **יישום:** התייחסות לרצף N וקטורי המאפיינים כאל "טוקנים" קלט, מה שמאפשר למנגנון הקשב העצמי לשקלל את חשיבותם של צעדי זמן ומאפיינים שונים. 12 מציע שימוש ברשת מבוססת-Transformer עבור רשת ה-actor בסוכן PPO למיטוב תיקים. 13 משתמש ב-LLM (המבוסס על טרנספורמר) כרשת המדיניות.
* **אוטואנקודרים (AE) / אוטואנקודרים וריאציוניים (VAE):**
  + **יתרונות:** למידה בלתי מונחית של ייצוגים דחוסים. VAEs מספקים מרחב סמוי (latent space) הסתברותי.1 יכולים לשמש לחילוץ מאפיינים והפחתת ממדיות. אוטואנקודרים גנרטיביים במסגרת DERL 1 (הפניה 28) לומדים הטמעות הלוכדות דינמיקות מעבר.
  + **חסרונות:** אוטואנקודרים סטנדרטיים עשויים שלא לעבור אופטימיזציה מפורשת עבור יכולת חיזוי ספציפית למשימה, אלא אם כן הם חלק ממערכת מקצה-לקצה. VAEs יכולים להיות מורכבים לאימון.
  + **יישום:** אימון AE/VAE על מטריצות הקלט N×F כדי ללמוד הטמעה קומפקטית משכבת צוואר הבקבוק. אם מאומנים מקצה-לקצה עם RL (כמו ב-DERL), הם לומדים הטמעות רלוונטיות למשימה.
* **שילוב ארכיטקטורות:** מודלים היברידיים (למשל, CNN-LSTM 1, LSTM-GNN 1, VAE-Transformer-LSTM 1) נפוצים כדי למנף את החוזקות של כל רכיב. לדוגמה, שכבות CNN לחילוץ מאפיינים ראשוני, ואחריהן LSTM/Transformer למודליזציה של הרצף.

בחירת ארכיטקטורת ההטמעה אינה בלתי תלויה באופי מטריצת הקלט (N×F). אם N גדול מאוד (רצפים ארוכים), טרנספורמרים עם קשב יעיל או LSTMs עשויים להיות מועדפים. אם דפוסים מקומיים בזמן חשובים מאוד, CNNs חזקים. ההטרוגניות של F המאפיינים משנה גם היא; ארכיטקטורות מסוימות עשויות להתמודד טוב יותר עם סוגי מאפיינים מגוונים לאחר נרמול ראשוני.1

שאיפתו של המשתמש לאמן את רשת ההטמעה באמצעות שגיאת TD מסוכן ה-RL (אימון מקצה-לקצה) משמעה שרשת ההטמעה אינה רק מעבד-מקדים קבוע. היא לומדת באופן פעיל ליצור ייצוגים שהם *שימושיים* לקבלת ההחלטות של סוכן ה-RL ולמקסום התגמול.1 הדבר הופך ארכיטקטורות כמו אלו שב-DERL 1 (הפניה 28) או DeepTrader 1 (הפניה 45) לרלוונטיות במיוחד. אימון מקצה-לקצה זה הופך את רשת ההטמעה לחלק אינטגרלי ממערכת ה-RL, הלומדת לייצר הטמעות המסייעות ישירות לסוכן ה-RL למקסם תגמולים.

טבלה 2.2: השוואת ארכיטקטורות רשת נוירונים להטמעת מטריצת קלט

(מבוסס על 1)

| **ארכיטקטורה** | **חוזק עיקרי להטמעת סדרות זמן פיננסיות** | **חולשות/אתגרים עיקריים** | **התאמה למטריצת N×F (מורכבות זמן ומאפיינים)** | **הערות על אימון מקצה-לקצה עם RL** |
| --- | --- | --- | --- | --- |
| 1D-CNN | חילוץ מאפיינים מקומיים ודפוסים זמניים קצרים, יעילות חישובית. | שדה קליטה מוגבל לתלויות ארוכות טווח (אלא אם משתמשים בקונבולוציות מורחבות). | טובה לרצפים קצרים-בינוניים (N), יכולה לטפל במספר מאפיינים (F) במקביל. | יכולה לשמש כשכבת חילוץ מאפיינים ראשונית לפני רשת RL עמוקה יותר; הגרדיאנטים יכולים לזרום דרכה. |
| LSTM/GRU | לכידת תלויות זמניות ארוכות-טווח והקשר ברצפים. | עיבוד סדרתי מגביל מקבול, עלולה להיות איטית לאימון, רגישות להיפר-פרמטרים. | מתאימה היטב לרצפים ארוכים (N), יכולה לעבד וקטורי מאפיינים (F) בכל צעד זמן. | סטנדרטית ב-RL מבוסס-רצף; המצב החבוי האחרון משמש לעיתים קרובות כהטמעה. |
| Transformer | לכידת תלויות גלובליות באמצעות קשב עצמי, מקבול גבוה. | דורשת נתונים רבים, יקרה חישובית לרצפים ארוכים מאוד (ללא גרסאות יעילות), כיול מורכב. | יכולה להתמודד עם רצפים ארוכים (N) ומספר רב של מאפיינים (F), במיוחד עם מנגנוני קשב יעילים. | פחות נפוצה כשכבת הטמעה ישירה ב-RL קלאסי בהשוואה ל-LSTM, אך אפשרית; הקשב יכול ללמוד אילו חלקים מההיסטוריה (N) והמאפיינים (F) רלוונטיים. |
| Autoencoder/VAE | למידה בלתי מונחית של ייצוגים דחוסים, הפחתת רעשים. VAEs לוכדים התפלגות סמויה. | הטמעה סטנדרטית אינה מותאמת למשימת RL ספציפית אלא אם מאומנת מקצה-לקצה. | יכולה לדחוס את כל מטריצת N×F או כל וקטור מאפיין F בנפרד. | אימון מקצה-לקצה (כמו ב-DERL 1) הופך את ההטמעה לרלוונטית למשימה. |
| היברידי (למשל, CNN+LSTM) | שילוב חוזקות (למשל, CNN לחילוץ מאפיינים מקומיים, LSTM למודליזציה של תלויות זמניות). | מורכבות מוגברת בתכנון ובאימון. | יכולה להתאים למגוון רחב של מאפייני N×F על ידי התאמת הרכיבים. | נפוץ; מאפשר לכל רכיב להתמחות במה שהוא טוב בו, תוך שהגרדיאנטים זורמים דרך כל המבנה. |

טבלה זו מספקת השוואה מובנית כדי לסייע למשתמש לבחור את ארכיטקטורת הרשת הנוירונית המתאימה ביותר לשכבת ההטמעה שלו. היא מקשרת בין בחירות ארכיטקטוניות לצרכים הספציפיים של נתוני סדרות זמן פיננסיות ולמטרה של אימון RL מקצה-לקצה.

## **חלק III: תכנון מערכת למידת חיזוק מרובת סוכנים (MARL)**

חלק זה מפרט את מערכת ה-MARL אשר צורכת את הייצוג הווקטורי המוטמע לצורך קבלת החלטות מסחר. הוא כולל את בחירת ארכיטקטורת הסוכנים, הגדרת מרחבי המצב והפעולה, עיצוב המדיניות המשותפת, והבטחת שיתוף פעולה.

### **סעיף 3.1: ארכיטקטורת MARL: הטרוגניות סוכנים ופרדיגמת אימון**

#### **3.1.1. בחינת השימוש בסוגי סוכני RL הטרוגניים**

המשתמש הביע עניין בבחינת השימוש בסוגי סוכני RL הטרוגניים (כגון Q-learning, DDPG) במערכת אחת (שאילתת משתמש). גישה זו, שבה סוכנים שונים עשויים להשתמש באלגוריתמי RL שונים או להיות בעלי תפקידים שונים, יכולה להוביל למערכת גמישה ובעלת יכולות מגוונות יותר.16

* **התמחות מבוססת נתונים/משימה:**
  + **Q-learning:** מתאים לסביבות עם מרחבי פעולה ומצב בדידים. יכול לשמש סוכן שאחראי על החלטות בדידות, כגון בחירת משטר מסחר כללי או הפעלת מודול ספציפי, בהינתן קלט שעבר דיסקרטיזציה.18
  + **DDPG (Deep Deterministic Policy Gradient):** אלגוריתם actor-critic המתאים למרחבי פעולה רציפים.18 יכול לשמש סוכן שאחראי על קביעת גודל פוזיציה רציף או רמות מחיר מדויקות לפקודות גבול, בהתבסס על קלט רציף.11
  + **PPO (Proximal Policy Optimization):** אלגוריתם actor-critic פופולרי היכול להתמודד עם מרחבי פעולה בדידים ורציפים כאחד, וידוע ביציבותו היחסית.21 יכול לשמש כאלגוריתם ברירת מחדל עבור סוכנים רבים או עבור סוכן מרכזי המקבל החלטות מורכבות.
* **יתרונות ההטרוגניות:**
  + **התאמה למאפייני הבעיה:** כל סוכן יכול להשתמש באלגוריתם המתאים ביותר למאפייני הנתונים שהוא מעבד ולסוג ההחלטות שהוא מקבל.16
  + **יכולות מגוונות:** שילוב סוגי סוכנים שונים מאפשר למערכת הכוללת להפגין מגוון רחב יותר של התנהגויות ויכולות הסתגלות.17
  + **מדרגיות פונקציונלית:** קל יותר להוסיף יכולות חדשות למערכת על ידי הוספת סוכנים חדשים עם התמחויות ספציפיות.16
* **אתגרים:**
  + **תיאום בין סוכנים הטרוגניים:** הבטחת שיתוף פעולה יעיל בין סוכנים המשתמשים באלגוריתמים שונים ובעלי פונקציות מטרה שונות עשויה להיות מאתגרת.18
  + **מורכבות אימון:** אימון מערכת עם סוכנים הטרוגניים דורש תכנון קפדני של סביבת האימון, פונקציות התגמול, ומנגנוני התקשורת (אם קיימים).23

בעוד שהרעיון של סוכנים הטרוגניים לחלוטין (כל אחד עם אלגוריתם RL שונה) הוא מורכב, גישה מעשית יותר עשויה להיות שימוש באלגוריתם MARL מרכזי (כמו MAPPO) עם מדיניות משותפת, אך לאפשר לסוכנים "התמחות" דרך האופן שבו הם מפרשים את הקלט שלהם או דרך תפקידים פונקציונליים שונים במערכת היררכית. לדוגמה, מודל ה-Hierarchical Reinforced Trader (HRT) משתמש ב-PPO עבור בקר ברמה גבוהה (בחירת מניות) וב-DDPG עבור בקר ברמה נמוכה (ביצוע עסקאות) 22, מה שמדגים הטרוגניות פונקציונלית.

#### **3.1.2. ארכיטקטורת MARL: אימון מרכזי עם ביצוע מבוזר (CTDE)**

* **פרדיגמת CTDE:** בהינתן הדרישה ל"שיתוף פעולה מושלם" בין הסוכנים ולמדיניות מסחר משותפת, פרדיגמת CTDE היא המתאימה ביותר.1
  + במהלך **האימון המרכזי**, ניתן לנצל מידע גלובלי – כגון מצבים ופעולות של כל הסוכנים, או פונקציית תגמול גלובלית – כדי לאמן את מדיניות הסוכנים או פונקציית ערך מרכזית (critic). הדבר מסייע בהתמודדות עם בעיית אי-הסטציונריות הנגרמת משינוי מדיניות של סוכנים אחרים.25
  + במהלך **הביצוע המבוזר**, כל סוכן פועל על סמך התצפית המקומית שלו (הנגזרת מהתצוגה Mi​ לאחר הטמעה) והמדיניות המשותפת שנלמדה.1
* **אלגוריתם MAPPO (Multi-Agent Proximal Policy Optimization):** אלגוריתם זה מהווה בחירה טבעית עבור ארכיטקטורת CTDE עם מדיניות משותפת, והוא הרחבה של אלגוריתם PPO הפופולרי לעולם מרובה הסוכנים.1 MAPPO מתאים במיוחד למשימות שיתופיות שבהן כל הסוכנים חולקים את אותה מדיניות ומאומנים באמצעות critic מרכזי המשתמש במידע גלובלי.
* **ספריות מימוש:** ספריות קוד פתוח כגון Ray RLlib ו-PettingZoo מספקות תשתית חזקה לפיתוח ואימון מערכות MARL מורכבות, ותומכות באלגוריתמים כמו MAPPO ובפרדיגמת CTDE.1 EpyMARL היא מסגרת נוספת המרחיבה את PyMARL ותומכת ב-MAPPO ובסביבות PettingZoo.29 MARLlib היא ספרייה נוספת המבוססת על Ray ו-RLlib, ומטרתה לפשט את הפיתוח של MARL.26

#### **3.1.3. עיצוב קלט הסוכנים ("תצוגות" Mi​) והטמעת נתונים**

* **חלוקה דינמית של נתונים לתצוגות:** במקום גישה של חלוקה אקראית וסטטית של הנתונים למטריצות נפרדות עבור כל סוכן, הגישה המועדפת היא חלוקה דינמית. מנוע זיהוי המשטרים (MMD) יזהה את מצב השוק הנוכחי, ועל סמך מצב זה, הנתונים הגולמיים (לאחר המרתם למאפיינים כפי שתואר בסעיף 1.2.2) יחולקו או יסוננו ליצירת "תצוגות" (Mi​) ייחודיות (או מותאמות) עבור כל סוכן או קבוצת סוכנים.1
  + **תוכן התצוגה Mi​:** כל תצוגה Mi​ תכיל רצף (N) של וקטורי מאפיינים (F) הרלוונטיים לאותו סוכן או לאותו היבט של השוק. זה יכול לכלול 1:
    - נתוני OHLCV רלוונטיים (5 דקות ו-30 דקות) מהמקטע/היבט שהוקצה לסוכן.
    - ערכי האינדיקטורים (MLMI, NW-RQK, FVG) המחושבים על בסיס נתונים אלו עבור המקטע.
    - מאפיינים כמותיים מפרופיל השוק (למשל, קרבה ל-LVN, "חוזק תחנת LVN" קרובה).
    - מאפיינים גיאומטריים של ה"צורה ההרמונית" האחרונה שזוהתה (זוויות, יחסי צלעות מנורמלים), אם רלוונטי למקטע הנתונים של הסוכן.
  + בנוסף, כל סוכן יקבל את פלט זיהוי המשטר הגלובלי ממנוע ה-MMD.
* **טיפול במאפיינים מבוססי אירועים:** כפי שנדון בסעיף 1.2.3, מאפיינים כמו צורות הרמוניות, שאינם מתקיימים בכל צעד זמן, ייוצגו בתצוגה Mi​ באמצעות ערכי ברירת מחדל, דגלים בינאריים, או מאפייני "זמן מאז האירוע".1
* **אורך רצף הקלט (N ב-N×F):** היפר-פרמטר קריטי שייקבע באמצעות כיול, תוך איזון בין הקשר היסטורי מספק למורכבות חישובית.1
* **הטמעת נתונים (Embedding) באמצעות רשתות נוירונים:**
  + **מטרה:** להמיר את התצוגה Mi​ (שהיא הטרוגנית ועשויה להיות בעלת ממדיות גבוהה) לייצוג וקטורי דחוס, אינפורמטיבי ואחיד. ייצוג זה ישמש כקלט (מצב si​) לרשת המדיניות של סוכן ה-RL.1
  + **אימון מקצה לקצה (End-to-End):** רשת ההטמעה תאומן כחלק אינטגרלי מארכיטקטורת הסוכן. הגרדיאנטים מאות השגיאה של סוכן ה-RL (למשל, TD error) יתפשטו לאחור דרך רשת המדיניות/ערך וגם דרך רשת ההטמעה, ויכוונו את רשת ההטמעה לייצר ייצוגים שימושיים למקסום התגמול.1
  + **ארכיטקטורות מועמדות:** Transformer נבחרה כארכיטקטורה מבטיחה בשל יכולתה ללכוד תלויות גלובליות ברצפים באמצעות מנגנוני קשב, והתאמתה לנתונים סדרתיים מורכבים.1 עם זאת, גם CNNs ו-LSTMs (או שילובים היברידיים) הם אפשרויות תקפות (ראה טבלה 2.2).
  + *ספריות למימוש רשתות הטמעה:* PyTorch, TensorFlow/Keras (כלליות ללמידה עמוקה).

הגישה של "חלוקה דינמית של נתונים לתצוגות" 1 היא מתוחכמת, אך מימושה דורש הגדרה ברורה של הלוגיקה שבאמצעותה פלט ה-MMD מתורגם לסינון או ארגון מחדש של המאפיינים עבור כל "תצוגה". האם כל סוכן יקבל תת-קבוצה שונה של *סוגי* מאפיינים, או שכל הסוכנים יקבלו את כל סוגי המאפיינים אך עם *דגשים* שונים (למשל, חלונות זמן שונים, רמות פירוט שונות) בהתאם למשטר? שאלה זו קריטית, שכן אם החלוקה/סינון אינם מוגדרים היטב, התצוגות עלולות להיות לא אינפורמטיביות או סותרות, גם אם המדיניות משותפת.1

ההחלטה לאמן את רשת ההטמעה מקצה לקצה עם סוכן ה-RL 1 היא חזקה, אך עלולה להפוך את תהליך האימון למורכב ואיטי יותר. יש לשקול אפשרות של אימון מקדים (pre-training) של רשת ההטמעה במשימה בלתי מונחית (למשל, אוטואנקודר על נתוני שוק) כדי לאתחל אותה עם משקולות טובות יותר לפני האימון המשותף עם ה-RL.14 גישה זו עשויה לשפר את יציבות ומהירות האימון של מערכת ה-RL, אך ההטמעה המקדימה עשויה לא להיות אופטימלית למשימת ה-RL הספציפית.

#### **3.1.4. מדיניות משותפת ומרחב פעולות**

* **מדיניות משותפת (π):** כל הסוכנים במערכת יפעלו על פי אותה מדיניות מסחר. מדיניות זו תמפה את המצב המוטמע של הסוכן (הנגזר מהתצוגה Mi​ שלו) לפעולה או להתפלגות על פני פעולות.1
* **פרשנות הסינרגיות:** כפי שהובהר ב-1, הסוכן אינו "בוחר" איזו מארבע הסינרגיות להפעיל. השוק, באמצעות צירוף האותות, קובע איזו סינרגיה מתממשת. המדיניות המשותפת צריכה ללמוד כיצד לפעול בצורה אופטימלית (כניסה, גודל פוזיציה, ניהול סיכונים) *בהינתן* שהתממשה סינרגיה מסוימת, כפי שהיא משתקפת במאפייני המשולש ההרמוני ובשאר רכיבי המצב.
* **מרחב פעולות היברידי:** מרחב הפעולות של הסוכנים ישקף את הפעולות האפשריות במסחר בחוזים עתידיים וישלב רכיבים בדידים ורציפים 1:
  + **כיוון עסקה (בדיד):** {קנייה (Long), מכירה (Short), החזקה/ללא פעולה}.
  + **סוג פקודה (בדיד):** {פקודת שוק (Market), פקודת גבול (Limit)}.
  + **גודל הפקודה (רציף/בדיד מוגדר):** מספר החוזים. רכיב זה צריך להיות דינמי ולהיות מושפע, בין היתר, מהערכת "חוזק תחנת ה-LVN" ו"איכות הצורה ההרמונית".
  + **מחיר פקודת גבול (רציף):** אם נבחרה פקודת גבול, יש לקבוע את רמת המחיר.
  + **פעולות ניהול נוספות (בדיד, אופציונלי):** {ביטול פקודה קיימת, שינוי פקודה קיימת}.
  + אלגוריתמים מודרניים של RL כמו PPO ו-SAC יכולים להתמודד עם מרחבי פעולה היברידיים, לעיתים באמצעות "ראשים" נפרדים ברשת המדיניות.1

#### **3.1.5. הבטחת "שיתוף פעולה מושלם"**

המונח "שיתוף פעולה מושלם" מחייב שהפעולות של כל הסוכנים, על אף שהן מבוססות על תצוגות חלקיות ושונות, יהיו קוהרנטיות ויתרמו למטרת המערכת הכוללת.1

* **מדיניות משותפת ופונקציית תגמול גלובלית:** הדרך הישירה ביותר להשגת תיאום היא באמצעות אימון המדיניות המשותפת עם פונקציית תגמול גלובלית. ההנחה היא שהסוכנים ילמדו באופן אימפליציטי לתאם את פעולותיהם כדי למקסם את התגמול המשותף.1
* **חשיבות ההקשר הגלובלי (מפלט ה-MMD):** מתן מידע על משטר השוק הנוכחי לכל הסוכנים הוא קריטי, שכן הוא מספק נקודת ייחוס משותפת ומסייע למדיניות המשותפת להתאים את פעולותיה להקשר הרחב יותר.1
* **אתגר הנצפות החלקית:** הצלחת גישת ה-MARL עם מדיניות משותפת תלויה ביכולת המדיניות להכליל היטב על פני התצוגות החלקיות השונות (Mi​) ועדיין לייצר פעולות קוהרנטיות ברמה הגלובלית. אם יתברר שהתצפיות המקומיות אינן מספקות די מידע, ניתן לשקול מנגנוני תקשורת מפורשים בין סוכנים (למשל, CommNet, IC3Net 1) או פרוטוקולי החלטה קולקטיביים, אם כי אלו מוסיפים מורכבות.

טבלה 3.1: הגדרת מרחב המצב (Mi​) ומרחב הפעולות של סוכני MARL

(מבוסס על 1)

| **רכיב** | **תיאור / מאפיינים** | **הערות** |
| --- | --- | --- |
| **מרחב המצב (תצוגה Mi​ לכל סוכן)** | וקטור מאפיינים (או רצף וקטורים N×F) המורכב מ: | כל סוכן מקבל Mi​ ייחודי (או מותאם דינמית) + פרמטרי משטר גלובליים. |
|  | 1. נתוני OHLCV/הייקן אשי (5ד', 30ד') רלוונטיים למקטע/היבט של הסוכן. |  |
|  | 2. ערכי אינדיקטורים (MLMI, NW-RQK, FVG) מחושבים על המקטע/היבט. |  |
|  | 3. מאפיינים גיאומטריים של "צורה הרמונית" אחרונה (זוויות, יחסי צלעות מנורמלים), אם רלוונטי. | נגזר משלושת האותות. |
|  | 4. מאפייני "חוזק תחנת LVN" קרובה (למשל, מספר בדיקות היסטוריות, נפח בבדיקות), אם רלוונטי. | נגזר מפרופיל שוק. |
|  | 5. פלט מנוע זיהוי המשטרים (MMD) – גלובלי לכל הסוכנים. | מספק הקשר מאקרו. |
|  | 6. (אופציונלי) מידע על פוזיציות פתוחות של הסוכן/המערכת. |  |
| **מרחב הפעולות (משותף לכל הסוכנים)** | מרחב היברידי (בדיד ורציף): |  |
|  | 1. **כיוון עסקה (בדיד):** {קנייה (Long), מכירה (Short), החזקה/ללא פעולה}. |  |
|  | 2. **סוג פקודה (בדיד):** {פקודת שוק (Market), פקודת גבול (Limit)}. |  |
|  | 3. **גודל הפקודה (רציף/בדיד מוגדר):** מספר החוזים. | יושפע מ"חוזק LVN" ו"איכות הצורה ההרמונית". |
|  | 4. **מחיר פקודת גבול (רציף):** אם נבחרה פקודת גבול (למשל, יחסית למחיר נוכחי או LVN). |  |
|  | 5. **(אופציונלי) פעולות ניהול נוספות (בדיד):** {ביטול פקודה קיימת, שינוי פקודה קיימת}. |  |

טבלה זו מגדירה את "שפת התקשורת" בין הסוכנים לסביבה. היא קריטית להבנת המידע העומד לרשות הסוכנים לצורך קבלת החלטות, והאפשרויות העומדות בפניהם לפעולה. פירוט זה חיוני למימוש הסביבה והסוכנים. אם המצב אינו מוגדר היטב (למשל, חסר מידע קריטי או כולל יותר מדי רעש), או אם מרחב הפעולות אינו מאפשר לסוכן לבצע את הפעולות הנדרשות, יכולת הלמידה של הסוכן תיפגע קשות.

#### **3.1.6. הנדסת פונקציית תגמול למסחר שיתופי**

פונקציית התגמול (R) היא קריטית בלמידת חיזוק, שכן היא מנחה את תהליך הלמידה של הסוכנים. במערכת MARL שיתופית עם מדיניות משותפת, פונקציית התגמול צריכה להיות גלובלית ולשקף את מטרות הצוות הכוללות.1

* **איזון בין רווח והפסד, תשואה מותאמת-סיכון וביצוע סינרגיות:**
  + **רווח והפסד של התיק (P&L):** מדד ביצועים מרכזי, אך יש לאזנו עם רכיבי סיכון.1
  + **תשואה מותאמת-סיכון:** יש לשלב מדדים כמו יחס שארפ (Sharpe Ratio) או יחס סורטינו (Sortino Ratio), המחושבים על חלון מתגלגל. מדדים אלו מעודדים רווחיות תוך ניהול סיכונים.1
  + **ציון ביצוע סינרגיה:** יש לתגמל את המערכת על זיהוי וביצוע מוצלח של כל אחת מארבע הסינרגיות המוגדרות, כאשר תנאי השוק מתאימים. הדבר יבטיח שהסוכנים לומדים ליישם את כל רכיבי האסטרטגיה המשותפת. התגמול יכול להיות בינארי או רציף.1
  + **עלויות עסקה:** יש להכליל עמלות עסקה, החלקה (slippage) צפויה, ועלויות החזקת פוזיציה, כדי להבטיח שהאסטרטגיה רווחית גם לאחר התחשבות בעלויות מסחר ריאליות.1
* **עונשים על משיכות גדולות וחוסר תיאום:**
  + **עונשים על סיכון:** יש להטיל עונשים (תגמול שלילי) על משיכות גדולות (maximum drawdown), על תנודתיות גבוהה של התיק, או על חריגה ממגבלות סיכון שהוגדרו מראש.1
  + **עונש על חוסר תיאום (אופציונלי):** אם ניתן לזהות מצבים שבהם פעולות הסוכנים סותרות ברמה הגלובלית, ניתן לשקול הוספת רכיב עונש. עם זאת, הגדרה אובייקטיבית של "חוסר תיאום" במצב כזה היא מאתגרת.1
* **אסטרטגיות להקצאת אשראי (במידת הצורך):**
  + במערכת MARL שיתופית עם תגמול גלובלי, בעיית הקצאת האשראי (credit assignment) – כיצד לקבוע את תרומתו של כל סוכן בודד להצלחה או לכישלון הכולל – היא מורכבת.1 אם כל הסוכנים חולקים את אותה מדיניות, ייתכן שאין צורך בהקצאת אשראי מפורשת.
  + אם יידרש, ניתן לשקול שיטות כמו Value Decomposition Networks (VDN) ו-QMIX 1, Counterfactual Rewards 35, או Shapley Values 1, אם כי האחרון יקר חישובית.
  + המשתמש העלה שאלה לגבי תזמון הטמעת מנגנוני הקצאת אשראי.1 סביר שטכניקות אלו, אם יידרשו, יהיו רלוונטיות יותר בשלבים מתקדמים של אופטימיזציית המערכת.

שיקול נוסף הוא כיצד לשלב בתגמול את הרעיון ש"צורות הרמוניות טובות" או "תחנות LVN חזקות" אמורות להוביל לתוצאות טובות יותר.1 נראה שעדיף שכניסות איכותיות יותר יובילו באופן עקיף לרווח גבוה יותר או ליחס סיכון/סיכוי טוב יותר, ובכך יתוגמלו דרך הרכיבים האחרים של הפונקציה, כדי לא לסבך יתר על המידה את פונקציית התגמול.

טבלה 3.2: רכיבי פונקציית התגמול הגלובלית וניסוחם

(מבוסס על 1)

| **רכיב תגמול** | **ניסוח מתמטי / לוגי (רעיוני)** | **רציונל להכללה** | **משקל / חשיבות פוטנציאליים** |
| --- | --- | --- | --- |
| רווח והפסד (P&L) של התיק (מבוסס צוות) | ΔPortfolioValuet​ | מדד ביצועים מרכזי, משקף את המטרה העיקרית של מסחר. | גבוה, אך מאוזן עם רכיבי סיכון. |
| תשואה מותאמת לסיכון | יחס שארפ או סורטינו, מחושב על חלון מתגלגל. RS​=(Rp​−Rf​)/σp​ | מעודד רווחיות תוך ניהול סיכונים. | גבוה. |
| ציון ביצוע סינרגיה (לכל אחת מ-4 הסינרגיות) | למשל, +קבוע אם סינרגיה k זוהתה ובוצעה בהצלחה בתנאים מתאימים, 0 אחרת. או תגמול רציף המבוסס על איכות ביצוע הסינרגיה. | מבטיח שהסוכנים לומדים לבצע את כל רכיבי האסטרטגיה המשותפת. | בינוני-גבוה, תלוי בחשיבות כל סינרגיה. |
| עונש על משיכות גדולות (Max Drawdown Penalty) | −C×(CurrentDrawdown)2 אם Drawdown > סף | מניעת לקיחת סיכונים מוגזמים המובילים להפסדים גדולים. | בינוני-גבוה. |
| עונשי עלויות עסקה | סכום עמלות והחלקה (slippage) בפועל או מוערך. | מבטיח שהאסטרטגיה רווחית נטו, לאחר התחשבות בעלויות מסחר ריאליות. | גבוה. |
| (אופציונלי) מדד שיתוף פעולה / קוהרנטיות | עונש על פעולות סותרות בין סוכנים (אם ניתן לזיהוי ברמה הגלובלית), או תגמול על פעולות מתואמות התומכות במטרה גלובלית. | חיוני להשגת "שיתוף פעולה מושלם" אם המדיניות המשותפת לבדה אינה מספיקה. | בינוני. |
| (אופציונלי) רכיב נגד התאמת יתר | עונש על מורכבות מדיניות (למשל, נורמת L1/L2 של משקולות הרשת), תגמול על גיוון בפעולות (אם רצוי במסגרת הסופרפוזיציה). | מפחית סיכון ללמידת אסטרטגיות שבריריות המותאמות לרעש בנתוני האימון. | בינוני. |

פונקציית התגמול היא לב ליבה של למידת החיזוק, ועיצובה הקפדני חיוני להכוונת הסוכנים להתנהגות הרצויה. טבלה זו מאפשרת דיון מובנה על האיזונים השונים הנדרשים בין רווח, סיכון, ביצוע אסטרטגיה, והימנעות מהתאמת יתר. הסוכנים ילמדו לעשות כל מה שנדרש כדי למקסם את התגמול; לכן, אם התגמול אינו משקף נכונה את מטרות המערכת, התנהגות הסוכנים עלולה להיות תת-אופטימלית או אף מזיקה.

### **סעיף 3.2: חקירת עקרונות סופרפוזיציה בהשראת קוונטים לעמידות המערכת**

הרעיון של שילוב עקרונות "סופרפוזיציה" בהשראת מכניקת הקוונטים במערכת למידת חיזוק קלאסית הוא היבט חדשני ומאתגר של הפרויקט. מטרתו היא לבחון כיצד עקרונות אלו יכולים לתרום להגברת העמידות (robustness) והאמינות (reliability) של סוכני המסחר והמערכת כולה.1

#### **3.2.1. הבהרה: הדמיה קלאסית של עקרונות קוונטיים**

חשוב להדגיש, כפי שציין המשתמש 1, שהכוונה אינה לבנות מחשב קוונטי או להשתמש באלגוריתמים קוונטיים אמיתיים. במקום זאת, המטרה היא לשאוב השראה מרעיונות קונספטואליים ממכניקת הקוונטים – בעיקר מעיקרון הסופרפוזיציה – ולממש אנלוגיות קלאסיות שלהם.1 אלגוריתמים בהשראת קוונטים (QIAs) מדמים עקרונות קוונטיים באמצעות חומרה קלאסית, לרוב על ידי שימוש במודלים מתמטיים או בדגימה הסתברותית, והם נגישים יותר ליישום מיידי.1

#### **3.2.2. מטרה: שיפור עמידות ואמינות**

העיקרון המנחה מאחורי הדמיית סופרפוזיציה במודלים קלאסיים הוא היכולת לייצג ולעבד מספר מצבים או פעולות פוטנציאליים במקביל, לפני שהמערכת "מתחייבת" למצב או פעולה בודדים.1 גישה זו מאפשרת גמישות רבה יותר, חקירה יסודית יותר של מרחב הפתרונות, ובסופו של דבר, קבלת החלטות מושכלת יותר, במיוחד בתנאים של אי-ודאות גבוהה המאפיינים שווקים פיננסיים.1

#### **3.2.3. אסטרטגיות מימוש פרקטיות (קלאסיות)**

המשתמש הביע עניין לדחות את ההחלטה על המימוש הספציפי של סופרפוזיציה לשלב מאוחר יותר בפרויקט.1 עם זאת, חשוב לשרטט מספר אסטרטגיות קלאסיות אפשריות:

* **שיטות אנסמבל (Ensemble Methods):**
  + **רעיון:** אימון של מספר מדיניויות (או מספר גרסאות של המדיניות המשותפת). כל מדיניות באנסמבל יכולה לייצג "מצב" או "השערה" הנמצאת בסופרפוזיציה.1
  + **מימוש:** ההחלטה הסופית מתקבלת על ידי אגרגציה של הפעולות המוצעות (למשל, הצבעת רוב, ממוצע משוקלל).1 למידת חיזוק באנסמבל (ERL) הראתה שיפור ביכולת ההכללה ובעמידות.1
  + **ספריות:** scikit-learn (לאנסמבלים בסיסיים), וניתן לממש אנסמבלים של סוכני RL ב-RLlib.
* **בחירת פעולה הסתברותית (Probabilistic Action Selection):**
  + **רעיון:** המדיניות המשותפת מפיקה התפלגות הסתברות על פני כל הפעולות האפשריות.1
  + **מימוש:** הפעולה הסופית נדגמת מהתפלגות זו. גישה זו מייצגת את המצב ה"מורכב" (superposed) של שקילת מספר פעולות בו-זמנית.1
  + **ספריות:** PyTorch או TensorFlow מאפשרות מימוש רשתות מדיניות עם פלט softmax ליצירת התפלגויות. TorchRL מספקת מחלקות ייעודיות לשחקנים הסתברותיים.15
* **מנגנון דמוי-QIDDM להתחייבות מאוחרת להפעלת סינרגיה:**
  + **רעיון:** גישה זו, הנראית כמתאימה ביותר לרעיון המקורי של המשתמש 1, מבוססת על אנלוגיה ל"קריסת פונקציית הגל" הקוונטית.1
  + **מימוש:** הסוכנים מעריכים במקביל את התועלת או ההתאמה של הפעלת כל אחת מארבע הסינרגיות המוגדרות. "קריסה" לבחירה בסינרגיה ספציפית מתרחשת רק כאשר מתקיים "טריגר קריסה" מוגדר (למשל, אות חזק, אישור ממשטר MMD, קונצנזוס בין סוכנים, או ירידת אנטרופיה מתחת לסף).1
  + **ספריות:** אין ספרייה ייעודית ל-QIDDM, אך ניתן לממש את הלוגיקה באמצעות Numpy וספריות RL סטנדרטיות לניהול מצבים והסתברויות.

#### **3.2.4. אינטראקציה עם מסגרת ה-MARL**

שילוב טכניקות בהשראת סופרפוזיציה ישנה את הדינמיקה של מערכת ה-MARL. במקום ש-N סוכנים יפעלו כל אחד עם אותה מדיניות דטרמיניסטית, ייתכן שיהיו N סוכנים שכל אחד מהם מריץ פנימית אנסמבל של M מדיניויות, או ש-N הסוכנים עצמם, כמערכת, מהווים אנסמבל של החלטות פוטנציאליות.1 הדבר מגדיל מורכבות חישובית אך טומן פוטנציאל לשיפור עמידות. "סופרפוזיציה" הופכת למנגנון מובנה לחקירה וניצול של מגוון אסטרטגיות.1

ההערכה של מערכת כזו תדרוש מדדים הלוכדים לא רק רווח והפסד, אלא גם את גיוון החקירה, יציבות הביצועים בתנאי שוק משתנים, ויכולת המערכת להסתגל בצורה "חיננית" לזעזועי שוק בלתי צפויים.1 אתגר הקצאת האשראי עשוי להסתבך עוד יותר אם סוכן בודד מריץ אנסמבל פנימי של מדיניויות.1

טבלה 3.3: אפשרויות מימוש לסופרפוזיציה בהשראת קוונטים

(מבוסס על 1)

| **גישה** | **רעיון "בהשראת קוונטים" מרכזי** | **מנגנון לעמידות/אמינות** | **תרחיש מימוש מעשי למסחר MARL** | **יתרונות** | **חסרונות/אתגרים** |
| --- | --- | --- | --- | --- | --- |
| ביצוע מדיניות באנסמבל | חקירה מקבילה של "מצבי" מדיניות מרובים. | גיוון אסטרטגי; אם מדיניות אחת נכשלת, אחרות עשויות להצליח. אגרגציית החלטות יכולה לסנן רעשים. | כל אחד מ-N הסוכנים (או המערכת כולה) מריץ K גרסאות של המדיניות המשותפת. החלטת המסחר הסופית מתקבלת על ידי הצבעת רוב או ממוצע משוקלל של המלצות ה-K מדיניויות. | שיפור הכללה, עמידות לרעש, הפחתת התאמת יתר למודל יחיד. | מורכבות חישובית מוגברת, צורך בניהול מספר מודלים, סיכון להתאמת יתר של האנסמבל כולו אם כל המודלים דומים מדי. |
| בחירת פעולה הסתברותית | ייצוג מצב קוונטי כהתפלגות על פני מצבי בסיס. | הימנעות מהתחייבות מוקדמת לפעולה בודדת; מאפשר תגובה גמישה יותר לאי-ודאות. | רשת המדיניות המשותפת מפיקה התפלגות הסתברות על פני מרחב הפעולות האפשרי. הפעולה הסופית נדגמת מהתפלגות זו, או שנבחרת הפעולה בעלת ההסתברות הגבוהה ביותר לאחר שעברה סף ביטחון מסוים. | חקירה טבעית, גמישות בתגובה. | דורש כיול של רמת האקראיות (טמפרטורה בדגימה), עלול להוביל לחוסר החלטיות אם ההתפלגות שטוחה מדי. |
| התחייבות מאוחרת דמוית-QIDDM | אנלוגיה ל"קריסת פונקציית הגל" הקוונטית. | החלטות מתקבלות רק כאשר יש מספיק "ראיות" או ביטחון, מה שמפחית פעולות פזיזות בסביבות רועשות. | הסוכנים מעריכים במקביל את התועלת של הפעלת כל אחת מארבע הסינרגיות. "קריסה" לסינרגיה ספציפית מתרחשת כאשר תנאי שוק מסוימים (אות חזק, אישור MMD, קונצנזוס) עוברים סף קריטי. | החלטות מבוססות יותר, הפחתת מסחר יתר בתנאי אי-ודאות. | מורכבות בהגדרת "טריגרי הקריסה" והספים, עלול להוביל להחמצת הזדמנויות אם הספים גבוהים מדי או אם תהליך הערכת הסינרגיות איטי. |

טבלה זו נועדה לסייע בקבלת החלטה מושכלת לגבי אסטרטגיית הסופרפוזיציה המתאימה ביותר למערכת, בהתחשב ביתרונות, בחסרונות ובמורכבות המימוש של כל גישה.

## **חלק IV: ארכיטקטורות אימון, בדיקה ופריסה**

חלק זה מתמקד בארכיטקטורות ובתהליכים הנדרשים לאימון אפקטיבי של סוכני ה-MARL, לבדיקה קפדנית של האסטרטגיה, ולהבטחת מעבר חלק ואמין ממסגרות בדיקה למסחר חי.

### **סעיף 4.1: ארכיטקטורות פיתוח המבטיחות מעבר חלק מבדיקות למסחר חי**

האתגר של הבטחת עקביות לוגית בין סביבת הבדיקות לאחור (backtesting) לסביבת המסחר החי (live trading) הוא קריטי, במיוחד עבור מערכות MARL מורכבות עם רכיבים דינמיים כמו זיהוי משטרי MMD והנדסת מאפיינים מבוססת-אירועים (צורות הרמוניות).34

#### **4.1.1. עקרונות של ארכיטקטורה מונחית-אירועים (Event-Driven Architecture - EDA)**

* **עקרון ליבה:** זרימת המערכת נקבעת על ידי אירועים כגון עדכוני נתוני שוק, ביצועי פקודות, או יצירת אותות.40 גישה זו מאפשרת תגובה בזמן אמת וקבלת החלטות מהירה.40
* **רכיבים מרכזיים ב-EDA למסחר:**
  + **מפיקי אירועים (Event Producers):** הזנות נתוני שוק (למשל, מ-Rithmic), מקורות חדשות (אם רלוונטי), ממשקי API חיצוניים.40
  + **צרכני אירועים (Event Consumers):** אלגוריתמי המסחר (סוכני MARL), מודולי ניהול סיכונים, מנגנוני התראה.40
  + **ערוצי אירועים (Event Channels):** מערכות העברת הודעות (messaging systems) כגון Apache Kafka או RabbitMQ להעברת אירועים בין מפיקים לצרכנים.1
  + **מעבדי אירועים (Event Processors):** לוגיקה לעיבוד אירועים, כגון מסננים, טרנספורמטורים, ואגרגטורים.40
* **יתרונות EDA למעבר חלק:**
  + **עקביות לוגית:** שימוש באותה לוגיקת טיפול באירועים, חישוב מאפיינים, ומודול קבלת החלטות של הסוכן הן בסביבת הבדיקות (העוטפת את מנוע הבקטסט) והן במנוע המסחר החי, מבטיח שהסוכן מקבל החלטות על בסיס אותו זרם עיבוד מידע.39
  + **מודולריות:** EDA מאפשרת פיתוח ובדיקה של רכיבים נפרדים (למשל, מנוע MMD, מודול חישוב צורות הרמוניות) כיחידות עצמאיות המתקשרות דרך אירועים.
* **ספריות רלוונטיות ל-EDA בפייתון:**
  + kafka-python: לאינטגרציה עם Apache Kafka.40
  + pika: עבור RabbitMQ.
  + ספריות פייתון סטנדרטיות לטיפול בתורים ואירועים.

מערכת הבדיקות לאחור, במיוחד עבור מערכת מורכבת ודינמית זו, צריכה להיות מונחית-אירועים כדי לשקף נאמנה את תנאי המסחר החי. בעוד שמנועי בדיקה וקטוריים עשויים להיות מהירים יותר עבור אסטרטגיות פשוטות 43, הם פחות מתאימים למודלים של תלות בנתיב, החלטות מבוססות-אירועים (כמו היווצרות משולש הרמוני), ושינויי משטר דינמיים, שבהם סדר האירועים והתגובה המיידית אליהם הם קריטיים. שימוש באותה ליבת לוגיקה מונחית-אירועים לבדיקות ולמסחר חי הוא פרקטיקה מומלצת למזעור פערים.39

#### **4.1.2. בחירת מנוע בדיקות לאחור (Backtesting Engine)**

* **שיקולים לבחירה:**
  + **תמיכה ב-MARL:** היכולת לשלב סביבת PettingZoo או סביבת MARL מותאמת אישית.
  + **ארכיטקטורה מונחית-אירועים:** חיונית לעקביות עם מסחר חי (כפי שנדון לעיל).
  + **טיפול בנתוני חוזים עתידיים:** כולל גלגול חוזים, עלויות עסקה, החלקה.
  + **גמישות בהנדסת מאפיינים:** היכולת לשלב את הלוגיקה המורכבת של חישוב אינדיקטורים, LVN, וצורות הרמוניות.
  + **שפת תכנות:** פייתון היא השפה הדומיננטית בפרויקט זה.
* **אפשרויות קיימות (והתאמתן):**
  + **Backtrader:** פלטפורמה פופולרית בפייתון, מונחית-אירועים, גמישה מאוד אך דורשת מאמץ פיתוח משמעותי לשילוב MARL.44
  + **Zipline:** פותחה במקור על ידי Quantopian, מונחית-אירועים, אך פחות מתוחזקת כיום ועלולה לדרוש התאמות.43
  + **QuantConnect (מנוע LEAN):** מבוסס C# (עם מעטפת פייתון באמצעות Python.NET), מונחה-אירועים, מציע תשתית ענן אך עשוי להיות יקר ולא תומך באופן מיטבי בחוזים נצחיים של מטבעות דיגיטליים.39
  + **פיתוח מותאם אישית על בסיס PettingZoo:** האפשרות הסבירה ביותר עבור מערכת MARL מורכבת זו. סביבת PettingZoo תשמש כעטיפה למנוע סימולציית שוק מותאם אישית (שיכול לשאוב עקרונות ממנועי בקטסט קיימים או להיות מפותח מאפס), שיטפל בנתונים היסטוריים, ביצוע פקודות מדומה, וחישוב תגמולים. גישה זו מעניקה שליטה מלאה על הלוגיקה ומבטיחה התאמה לצרכים הייחודיים של האסטרטגיה.

#### **4.1.3. אופטימיזציית Walk-Forward**

* **חשיבות:** טכניקה רובסטית לבדיקה והתאמה של אסטרטגיות מסחר בסביבות משתנות, המדמה טוב יותר תנאי מסחר אמיתיים ומסייעת בהפחתת התאמת יתר.1
* **תהליך:**
  1. **חלוקת נתונים:** הנתונים ההיסטוריים מחולקים למקטעים עוקבים (חלונות אימון ובדיקה).
  2. **אופטימיזציה (In-Sample):** המערכת (כולל סוכני MARL והיפר-פרמטרים) מאומנת על מקטע נתונים היסטורי ראשון.
  3. **בדיקה (Out-of-Sample):** המערכת המאומנת נבדקת על מקטע הנתונים הבא מיד לאחר תקופת האימון.
  4. **הזזת החלון:** חלון האימון והבדיקה "זזים קדימה" בזמן, והתהליך חוזר על עצמו.
* **אתגרים במערכת MARL מורכבת:** אופטימיזציית Walk-Forward עבור מערכת עם רכיבים רבים (MMD, הטמעה, MARL, מודל עזר) והיפר-פרמטרים רבים היא תהליך יקר חישובית. יש להשתמש בכלים כמו Optuna ו-Ray Tune לאוטומציה של חיפוש היפר-פרמטרים במסגרת כל חלון אימון.1

אופטימיזציית Walk-Forward חיונית למערכת אדפטיבית זו, אך הפרמטרים ה"אופטימליים" הנמצאים בחלון Walk-Forward אחד עשויים שלא להיות אופטימליים בחלון הבא בשל אי-סטציונריות עמוקה או שינויי משטר שאינם נלכדים במלואם אפילו על ידי MMD. הדבר מרמז על צורך ביכולות למידה מתמשכת או הסתגלות מקוונת עבור סוכני ה-MARL מעבר לאימון תקופתי מחדש. שווקים פיננסיים הם מאוד לא-סטציונריים.1 MMD מנסה ללכוד זאת.1 עם זאת, ל-MMD עצמו יש פרמטרים וייתכן שהוא לא ילכוד את כל צורות אי-הסטציונריות או יזהה שינויים בעיכוב. מדיניות סוכן ה-MARL, שעברה אופטימיזציה באמצעות Walk-Forward, מבוססת על הנתונים שנראו בחלון האימון שלה. אם השוק עובר שבר מבני *בתוך* חלון בדיקה, או סוג של שינוי משטר שאינו מיוצג היטב במצבי ה-MMD, המדיניות ה"אופטימלית" הקודמת עלולה להתדרדר במהירות. הדבר מצביע על כך שהסתמכות בלעדית על אימון תקופתי מחדש באמצעות Walk-Forward עשויה שלא להספיק. סוכני ה-MARL עצמם עשויים להפיק תועלת מיכולות למידה מקוונת או מגישות מטא-למידה המאפשרות להם להתאים את מדיניותם במהירות רבה יותר למצבים חדשים שנתקלים בהם במהלך מסחר חי, אולי על ידי התאמת שיעורי למידה או אסטרטגיות חקירה בהתבסס על סחיפת ביצועים.

### **סעיף 4.2: סביבות ארגז חול (Sandbox) לאימון אפקטיבי של סוכני MARL**

אימון סוכני MARL, במיוחד עבור אסטרטגיות מסחר מורכבות, דורש סביבות אימון מבודדות, ניתנות לשחזור, ומדרגיות.

#### **4.2.1. תכנון ארכיטקטוני לאימון מבודד, מדרגי וניתן לשחזור**

* **רציונל:** סוכני RL, ובמיוחד MARL, זקוקים לסביבות מבודדות לניסויים מבלי להשפיע על מערכות מארחות, להבטחת שחזוריות של תוצאות, ולאפשר הרצות מקבילות (parallel rollouts) לאיסוף נתונים יעיל.45
* **ארכיטקטורה מנותקת (Decoupled Architecture):** הפרדה בין צמתי הסקה (inference nodes) (מבוססי GPU, עבור הסקת מדיניות LLM/כללית) לבין צמתי ביצוע ארגז חול (sandbox execution nodes) (מבוססי CPU,

#### Works cited

1. ניתוח מערכת מסחר אוטומטית.
2. How to Run Python Code 30 Sec before Every 5th Minute - Stack Overflow, accessed on June 8, 2025, <https://stackoverflow.com/questions/52539614/how-to-run-python-code-30-sec-before-every-5th-minute>
3. python-signal-processing/splearn/data/generate.py at main - GitHub, accessed on June 8, 2025, <https://github.com/jinglescode/python-signal-processing/blob/main/splearn/data/generate.py>
4. Numerical data: Normalization | Machine Learning - Google for Developers, accessed on June 8, 2025, <https://developers.google.com/machine-learning/crash-course/numerical-data/normalization>
5. Is it a good practice to always scale/normalize data for machine learning? - Cross Validated, accessed on June 8, 2025, <https://stats.stackexchange.com/questions/189652/is-it-a-good-practice-to-always-scale-normalize-data-for-machine-learning>
6. numpy.angle — NumPy v2.1 Manual, accessed on June 8, 2025, <https://numpy.org/doc/2.1/reference/generated/numpy.angle.html>
7. Find all angles of a given triangle - GeeksforGeeks, accessed on June 8, 2025, <https://www.geeksforgeeks.org/find-angles-given-triangle/>
8. 1D convolutional Neural Network architecture - PyTorch Forums, accessed on June 8, 2025, <https://discuss.pytorch.org/t/1d-convolutional-neural-network-architecture/67171>
9. Creating a Simple 1D CNN in PyTorch with Multiple Channels - Stack Overflow, accessed on June 8, 2025, <https://stackoverflow.com/questions/55720464/creating-a-simple-1d-cnn-in-pytorch-with-multiple-channels>
10. Reinforcement Learning Framework for Quantitative Trading - arXiv, accessed on June 8, 2025, <https://arxiv.org/html/2411.07585v1>
11. LSTM-DDPG for Trading with Variable Positions - PMC, accessed on June 8, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8512099/>
12. Adaptive Portfolio Optimization via PPO-HER: A Reinforcement Learning Framework for Non-Stationary Markets - ResearchGate, accessed on June 8, 2025, <https://www.researchgate.net/publication/391319997_Adaptive_Portfolio_Optimization_via_PPO-HER_A_Reinforcement_Learning_Framework_for_Non-Stationary_Markets>
13. FLAG-Trader: Fusion LLM-Agent with Gradient-based Reinforcement Learning for Financial Trading - arXiv, accessed on June 8, 2025, <https://arxiv.org/html/2502.11433v1>
14. Transfer Learning for Computer Vision Tutorial - PyTorch documentation, accessed on June 8, 2025, <https://docs.pytorch.org/tutorials/beginner/transfer_learning_tutorial.html>
15. torchrl.modules package - PyTorch, accessed on June 8, 2025, <https://pytorch.org/rl/main/reference/modules.html>
16. Your Guide to Mastering Heterogeneous Agents in Complex AI Systems - Number Analytics, accessed on June 8, 2025, <https://www.numberanalytics.com/blog/mastering-heterogeneous-agents-complex-ai-systems>
17. Heterogeneous Agents: Ultimate Guide to Multi‑Agent AI - Number Analytics, accessed on June 8, 2025, <https://www.numberanalytics.com/blog/heterogeneous-agents-ultimate-guide-multi-agent-ai>
18. A Review of Multi-Agent Reinforcement Learning Algorithms - MDPI, accessed on June 8, 2025, <https://www.mdpi.com/2079-9292/14/4/820>
19. (PDF) Hybrid QINN-BP for Algorithmic Trading: A Quantum-Inspired 5D Reinforcement Learning Approach for Financial Markets - ResearchGate, accessed on June 8, 2025, <https://www.researchgate.net/publication/389499352_Hybrid_QINN-BP_for_Algorithmic_Trading_A_Quantum-Inspired_5D_Reinforcement_Learning_Approach_for_Financial_Markets>
20. DDPGAgent - Keras-RL Documentation - Read the Docs, accessed on June 8, 2025, <https://keras-rl.readthedocs.io/en/latest/agents/ddpg/>
21. Encouraging Cooperation in Multi Agent Reinforcement Learning - Harvard DASH, accessed on June 8, 2025, <https://dash.harvard.edu/server/api/core/bitstreams/27cb52a1-7f0d-4d60-9cab-aee0e29d9208/content>
22. Hierarchical Reinforced Trader (HRT): A Bi-Level Approach for Optimizing Stock Selection and Execution - arXiv, accessed on June 8, 2025, <https://arxiv.org/html/2410.14927v1>
23. Kaleidoscope: Learnable Masks for Heterogeneous Multi-agent Reinforcement Learning - NIPS papers, accessed on June 8, 2025, <https://proceedings.neurips.cc/paper_files/paper/2024/file/274d0146144643ee2459a602123c60ff-Paper-Conference.pdf>
24. [2410.14927] Hierarchical Reinforced Trader (HRT): A Bi-Level Approach for Optimizing Stock Selection and Execution - arXiv, accessed on June 8, 2025, <https://arxiv.org/abs/2410.14927>
25. Fully Decentralized Cooperative Multi-Agent Reinforcement Learning: A Survey - arXiv, accessed on June 8, 2025, <https://arxiv.org/html/2401.04934v1>
26. MARLlib: A Scalable and Efficient Library For Multi-agent ..., accessed on June 8, 2025, <https://www.jmlr.org/papers/volume24/23-0378/23-0378.pdf>
27. [2506.02718] Heterogeneous Group-Based Reinforcement Learning for LLM-based Multi-Agent Systems - arXiv, accessed on June 8, 2025, <https://arxiv.org/abs/2506.02718>
28. A survey on multi-agent reinforcement learning and its application - ResearchGate, accessed on June 8, 2025, <https://www.researchgate.net/publication/378311162_A_survey_on_multi-agent_reinforcement_learning_and_its_application>
29. An Extended Benchmarking of Multi-Agent Reinforcement Learning Algorithms in Complex Fully Cooperative Tasks - arXiv, accessed on June 8, 2025, <https://arxiv.org/html/2502.04773v1>
30. Huey Song/epymarl - Gitee, accessed on June 8, 2025, <https://gitee.com/hueysong_sxu/epymarl?skip_mobile=true>
31. MAPS: Multi-Agent reinforcement learning-based Portfolio management System. - IJCAI, accessed on June 8, 2025, <https://www.ijcai.org/proceedings/2020/0623.pdf>
32. Towards Robust Multi-UAV Collaboration: MARL with Noise-Resilient Communication and Attention Mechanisms - arXiv, accessed on June 8, 2025, <https://arxiv.org/html/2503.02913v1>
33. Deep Reinforcement Learning in Trading - Quantra by QuantInsti, accessed on June 8, 2025, <https://quantra.quantinsti.com/course/deep-reinforcement-learning-trading>
34. Backtesting vs Live Trading: Key Factors for a Successful Algo Strategy, accessed on June 8, 2025, <https://www.hashstudioz.com/blog/backtesting-vs-live-trading-key-factors-for-a-successful-algo-strategy/>
35. Multi-Agent Reinforcement Learning (MARL) For Collaboration - HitechNectar, accessed on June 8, 2025, <https://hitechnectar.com/blogs/multi-agent-reinforcement-learning-when-machines-team-up/>
36. Multi-Agent Reinforcement Learning in AI | GeeksforGeeks, accessed on June 8, 2025, <https://www.geeksforgeeks.org/multi-agent-reinforcement-learning-in-ai/>
37. Nucleolus Credit Assignment for Effective Coalitions in Multi-agent Reinforcement Learning, accessed on June 8, 2025, <https://www.researchgate.net/publication/389548417_Nucleolus_Credit_Assignment_for_Effective_Coalitions_in_Multi-agent_Reinforcement_Learning>
38. Reinforcement Learning Tricks: Mastering AI With MARL, Epsilon-Greedy, And Reward Shaping - Yeti AI, accessed on June 8, 2025, <https://yetiai.com/reinforcement-learning-tricks/>
39. The problem of logical consistency between backtesting and live trading - Blueshift, accessed on June 8, 2025, <https://quantra.quantinsti.com/community/t/the-problem-of-logical-consistency-between-backtesting-and-live-trading/25335>
40. Event-Driven Architecture in Python for Trading - PyQuant News, accessed on June 8, 2025, <https://www.pyquantnews.com/free-python-resources/event-driven-architecture-in-python-for-trading>
41. How to Use Apache Kafka for Real-Time Data Streaming? - ProjectPro, accessed on June 8, 2025, <https://www.projectpro.io/article/kafka-for-real-time-streaming/916>
42. Creating a Data Science Pipeline for Real-Time Analytics Using Apache Kafka and Spark, accessed on June 8, 2025, <https://www.kdnuggets.com/creating-a-data-science-pipeline-for-real-time-analytics-using-apache-kafka-and-spark>
43. Why Backtests Run Fast or Slow: A Comparison of Zipline ..., accessed on June 8, 2025, <https://www.quantrocket.com/blog/backtest-speed-comparison/>
44. Backtrader Alternatives for Trading & Backtesting - Forex Tester, accessed on June 8, 2025, <https://forextester.com/blog/backtrader-alternatives>
45. Sandbox Infrastructure for Reinforcement Learning Agents - Daytona, accessed on June 8, 2025, <https://www.daytona.io/dotfiles/sandbox-infrastructure-for-reinforcement-learning-agents>
46. Why Don't We See Multi-Agent RL Trained in Large-Scale Open Worlds? - Reddit, accessed on June 8, 2025, <https://www.reddit.com/r/reinforcementlearning/comments/1jhcdnn/why_dont_we_see_multiagent_rl_trained_in/>