# **ארכיטקטורה מקיפה למערכת מסחר אלגוריתמית מרובת סוכנים (MARL) עבור חוזים עתידיים**

## **1. תקציר מנהלים**

דוח זה מציג תכנון ארכיטקטוני מקיף למערכת מסחר אלגוריתמית מתוחכמת, מבוססת למידת חיזוק מרובת סוכנים (MARL), המיועדת לפעול בשוק החוזים העתידיים. המערכת נשענת על ניתוח נתונים היסטוריים של חמש שנים ברזולוציות של 30 דקות ו-5 דקות, ומתמקדת בנכס בודד בכל פעם. הארכיטקטורה המוצעת משלבת מספר רכיבים חדשניים, ביניהם מערכת MARL "קטנה" לזיהוי שבעה מצבי שוק באמצעות Maximum Mean Discrepancy (MMD), מערכת MARL "גדולה" עם סוכנים ייעודיים לרכיבי מידע שונים הפועלים תחת מדיניות משותפת בפרדיגמת Centralized Training with Decentralized Execution (CTDE) ואלגוריתם Multi-Agent Proximal Policy Optimization (MAPPO), רכיב חיצוני לניתוח "צורות הרמוניות" (משולשים) מצירופי אותות לצורך ניהול סיכונים דינמי, ומנגנון "סופרפוזיציה" מבוסס קשב (Attention) לאגרגציית החלטות הסוכנים.

המלצות הליבה של הדוח כוללות שימוש ברשתות הטמעה (embedding) נפרדות ומותאמות לכל סוג סוכן בהתאם להטרוגניות הקלט, תוך אימון מקצה-לקצה עם מערכת ה-RL. באשר לשיתוף מידע בין סוכנים בזמן אמת, ההמלצה היא להסתמך בשלב ראשון על תיאום מרומז דרך פרדיגמת CTDE, מצב השוק המשותף (MMD) ומיזוג אותות על ידי רכיב ניתוח הצורות ההרמוניות, ולהימנע מתקשורת ישירה בין סוכנים אלא אם יוכח צורך מובהק. לתשתית הפריסה, עבור מסחר בנכס יחיד עם דרישות השהיה נמוכות לבורסה ספציפית, Windows VPS הממוקם בקרבה פיזית לשרתי הבורסה מועדף על פני פתרונות ענן גנריים. יש למנף מנועי הסקה מהירים כגון ONNX ו-TensorRT למסחר חי. שילוב מנגנוני Human-in-the-Loop (HITL) ואופטימיזציית היפר-פרמטרים באמצעות Optuna הם חיוניים לבניית מערכת אמינה וביצועית.

למערכת המוצעת פוטנציאל משמעותי לייצר אסטרטגיות מסחר אדפטיביות ועמידות, אך מימושה כרוך בהתמודדות עם אתגרים הקשורים למורכבותה, לדרישות האימון ולאימות קפדני של כל רכיביה. דוח זה מתווה נתיב לפיתוח מערכת זו, תוך התייחסות לשיקולים קריטיים הנוגעים למסחר חי.

## **2. אסטרטגיית מסחר יסודית וארכיטקטורת נתונים**

בטרם ניגש לתכנון מערכת ה-MARL, חיוני לנתח לעומק את הלוגיקה האסטרטגית העומדת בבסיסה. הבנה זו של רכיבי הליבה, האינדיקטורים, טווחי הזמן והאינטראקציות ביניהם, מהווה תנאי הכרחי לבניית מערכת אינטליגנטית שתדע לפרש את השוק ולפעול בו בצורה מיטבית.

### **2.1. אסטרטגיית ליבה מבוססת אינדיקטורים: MLMI, NW-RQK, FVG**

ליבת האסטרטגיה נשענת על שילוב אותות ממספר אינדיקטורים הפועלים על נרות Heiken Ashi (HA) ומספקים פרספקטיבות משלימות על דינמיקת השוק.

* **MLMI (Machine Learning Momentum Index):** אינדיקטור זה פועל על גרפים של 30 דקות של נרות HA.1 הוא משלב עקרונות מומנטום מסורתיים, המבוססים על Relative Strength Index (RSI) וממוצעים נעים משוקללים (WMA), עם אלגוריתם למידת מכונה מסוג k-Nearest Neighbors (k-NN) במטרה לייצר תחזיות מומנטום אדפטיביות.1 אות המסחר העיקרי נגזר בעת חצייה של קו התחזית של MLMI עם ה-WMA שלו.1 קיימת העדפה לשימוש בפרמטרי ברירת מחדל עבור MLMI (למשל, 'Prediction Data (k)', 'Trend length') לשמירה על רובסטיות בתנאי שוק משתנים.1
* **NW-RQK (Nadaraya-Watson Rational Quadratic Kernel Regression):** אינדיקטור זה פועל גם הוא על גרפים של 30 דקות של נרות HA.1 הוא משתמש ברגרסיית ליבה לא-פרמטרית עם ליבת Rational Quadratic, המקנה משקל גבוה יותר לנתונים עדכניים 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 במקורות, היא היותו "non-repainting" (אינו מצייר מחדש ערכים היסטוריים), תכונה הכרחית לאמינות תוצאות בדיקות לאחור (backtesting) ולביצועים עקביים במסחר חי.1 גם כאן, מועדפים פרמטרי ברירת מחדל.1
* **FVG (Fair Value Gap) - LuxAlgo:** אינדיקטור זה מזהה אזורים של חוסר איזון במחיר ופועל על גרפים של 5 דקות (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

ההעדפה לשימוש בפרמטרי ברירת מחדל עבור MLMI ו-NW-RQK 1 נובעת מהשאיפה לרובסטיות ולהימנעות מאופטימיזציית-יתר של פרמטרי האינדיקטורים עצמם. החלטה זו מעבירה את נטל ההסתגלות לתנאי שוק משתנים אל סוכני ה-MARL. משמעות הדבר היא שהסוכנים חייבים ללמוד לפרש את האותות ה"סטטיים" מהאינדיקטורים באופן שונה, בהתבסס על הקשר שוק רחב יותר. הדבר מחייב ייצוג מצב עשיר ואינפורמטיבי עבור סוכני ה-MARL, הכולל קלט אמין על מצב השוק (שיתקבל ממערכת ה-MMD). המדיניות הנלמדת על ידי סוכני ה-MARL "הגדולים" תצטרך להיות מתוחכמת מספיק כדי למפות צירוף של אותות אינדיקטור קבועים, מצב שוק והקשר נוסף לפעולות אופטימליות. בחירה עיצובית זו, המעדיפה יציבות מערכתית והפחתת התערבות ידנית ברמת האינדיקטור, מגדילה משמעותית את מורכבות הלמידה ואת דרישות הנתונים עבור רכיבי ה-MARL. הצלחת המערכת תלויה ביכולתם של סוכני ה-MARL לבצע פרשנות הקשרית זו.

השינוי במסגרת הזמן של ניתוח ה-FVG מדקה אחת 1 לחמש דקות 1 הוא בעל השלכות משמעותיות. מיטיגציית FVG מהווה קודקוד מרכזי ב"משולשים ההרמוניים" (שיתוארו בהמשך). שינוי מסגרת הזמן של FVG משנה באופן ישיר את הקואורדינטה הזמנית (ועל כן, גם את קואורדינטת המחיר, אם השוק נע) של קודקוד זה ביחס לאותות ה-MLMI וה-NW-RQK (הנגזרים מנתוני 30 דקות). כתוצאה מכך, המאפיינים הגיאומטריים (זוויות, אורכי צלעות, שטח) של "המשולשים ההרמוניים" ישתנו באופן יסודי. סוכני ה-MARL, האמורים ללמוד פרמטרי ניהול סיכונים ממאפיינים גיאומטריים אלו, יצטרכו כעת ללמוד מהתפלגות שונה של "צורות". הדבר מדגיש את הצורך של המערכת להיות עמידה לא רק לשינויי שוק חיצוניים אלא גם לשינויי תצורה פנימיים המשפיעים על מאפיינים נגזרים. תידרש הערכה מחודשת יסודית, ואולי אף אימון מחדש, של רכיב ניהול הסיכונים אם פרמטרים אסטרטגיים כאלה משתנים.

### **2.2. ניתוח פרופיל שוק: LVNs והערכת "חוזק תחנה"**

* **LVN (Low Volume Nodes):** צמתי נפח נמוך מזוהים מניתוח פרופיל שוק על נתוני 30 דקות. צמתים אלו מייצגים רמות מחיר עם נפח מסחר נמוך היסטורית ונחשבים ל"תחנות" משמעותיות או נקודות מפנה פוטנציאליות שבהן השוק עשוי להגיב.1
* **הערכת "חוזק תחנה":** חשיבותו של LVN מוערכת על בסיס אינטראקציות מחיר היסטוריות (מבחני תמיכה/התנגדות, נפח שנסחר ב-LVN במהלך בדיקות קודמות). "חוזק" זה ישפיע ישירות על החלטות גודל הפוזיציה שיתקבלו על ידי סוכני ה-MARL.1 ניתן להשתמש בספריית MarketProfile בפייתון לחישוב פרופיל שוק וזיהוי LVNs.4

הכנסת המושג "חוזק תחנת LVN" לאסטרטגיה מכניסה רכיב ניהול סיכונים דינמי. כימות "חוזק" זה באופן אובייקטיבי הוא משימת הנדסת מאפיינים מרכזית. סוכן ה-MARL יצטרך ללמוד את המתאם בין מדדי עוצמה מכומתים אלו (ומאפיינים הקשריים אחרים) לבין גודל הפוזיציה האופטימלי. לשם כך, יש לפתח שיטה אלגוריתמית שיטתית לכימות אינטראקציות היסטוריות אלו ל"ציון עוצמה" מספרי או למערך מאפיינים המייצגים עוצמה. מאפייני עוצמת LVN אלו יהפכו לחלק ממצב הקלט של סוכן ה-RL. רשת המדיניות של הסוכן תלמד מיפוי ממצב זה (הכולל עוצמת LVN, מאפייני משולש הרמוני, מצב שוק וכו') לגודל פוזיציה אופטימלי. הדבר הופך רעיון מסחר איכותני ("רמה חזקה") לקלט מכומת עבור פונקציית ניהול סיכונים דינמית ונלמדת, צעד משמעותי לקראת מסחר מבוסס בינה מלאכותית בעל ניואנסים.

### **2.3. עיבוד נתונים רב-ממדי והמרת Heiken Ashi**

* **מקורות נתונים:** המערכת תשתמש בנתוני OHLCV עבור חוזים עתידיים. ספקים אפשריים שהוזכרו כוללים את Rithmic 1, Databento 1 ו-dxFeed.1 השאילתה מציינת שימוש בחמש שנות נתונים היסטוריים ברזולוציות של 30 דקות ו-5 דקות.
* **המרת Heiken Ashi (HA):** כל נתוני ה-OHLCV הגולמיים (30 דקות ו-5 דקות) יומרו לנרות HA.1 הרציונל הוא החלקת מגמות, הפחתת רעשים והתאמה טובה יותר לאסטרטגיה. נוסחאות חישוב HA מפורטות ב-.1
* **ניקוי וסנכרון נתונים:** ייושמו נהלים סטנדרטיים לטיפול בנתונים חסרים, חריגות, והבטחת סנכרון חותמות זמן מדויק בין זרמי הנתונים.1
* **חוזים מתגלגלים:** תיושם לוגיקה לטיפול בגלגול חוזים עתידיים ליצירת סדרות נתונים רציפות.1

ההחלטה להמיר את *כל* נתוני המחיר ל-Heiken Ashi *לפני* חישוב אינדיקטורים כמו FVG ופרופיל שוק (עבור LVN) היא משמעותית.1 דפוסי FVG סטנדרטיים ופרופיל שוק קלאסי מוגדרים בדרך כלל על בסיס ערכי OHLC גולמיים. נרות HA, בשל אופי המיצוע שלהם, משנים ערכים אלו. כתוצאה מכך, חישוב FVG ו-LVN על נתוני HA יניב דפוסים ורמות שונים מאלו שהיו מתקבלים מנתונים גולמיים. המערכת תצטרך להתבסס על הגדרות FVG/LVN מותאמות אישית לנתוני HA, או לחשב FVG/LVN על נתונים גולמיים וליישר אותות/רמות אלו עם האותות מבוססי HA (MLMI/NW-RQK/MMD). לבחירה זו השלכות עמוקות על הנדסת המאפיינים. התכנון הנוכחי 1 נוטה לשימוש ב-HA עבור כל הרכיבים, מה שאומר שרכיבי ה-FVG וה-LVN יהיו בהכרח לא סטנדרטיים וידרשו אימות והתאמה ייחודיים.

### **2.4. ניתוח צורות הרמוניות (משולשים) לניהול סיכונים דינמי**

* **תפיסה:** צירוף שלושת האותות העיקריים (MLMI, NW-RQK, ומיטיגציית FVG) יוצר קודקודים של "משולש הרמוני" על גרף ה-5 דקות.1
* **מיצוי מאפיינים גיאומטריים:** מאפיינים כמותיים כגון זוויות, אורכי צלעות, יחסים ביניהם, ושטח המשולש מופקים מצורות אלו.1
* **מטרה:** מאפיינים גיאומטריים אלו משמשים כמדד ל"איכות" או "תוקף" של מערך הכניסה הפוטנציאלי, ומשפיעים ישירות על החלטות ניהול סיכונים (גודל פוזיציה, Stop-Loss/Take-Profit) המתקבלות על ידי סוכני ה-MARL.1 מידע כללי על זיהוי תבניות משולש וחישובים גיאומטריים ניתן למצוא ב-.10
* **נרמול:** מאפיינים גיאומטריים (במיוחד אורכים ושטח) חייבים לעבור נרמול (למשל, על ידי ATR) כדי להיות ברי-השוואה בתנאי שוק שונים.1
* **אתגר הקרנת אותות:** אותות MLMI/NW-RQK (מנתוני HA של 30 דקות) דורשים הקרנה מדויקת על גרף ה-5 דקות כדי להגדיר את קודקודי המשולש. בחירת נר ה-5 דקות הספציפי בתוך בר ה-30 דקות המייצג את האות משפיעה באופן משמעותי על הגיאומטריה.1 ניתן למצוא דיונים כלליים על הקרנת נתונים בין מסגרות זמן ב-.14

ההנחה המרכזית היא שתצורות גיאומטריות ספציפיות של משולשים אלו מתואמות עם הסתברות גבוהה יותר להצלחת העסקה או ליחס סיכון/תגמול טוב יותר.1 סוכני ה-MARL ילמדו להתאים את הסיכון על סמך מאפיינים גיאומטריים אלו. אם למאפיינים אלו אין כוח ניבוי אמיתי, הם עלולים להכניס רעש למערכת או לגרום לסוכנים ללמוד קורלציות מזויפות. לכן, אימות סטטיסטי קפדני של כוח הניבוי של מאפיינים גיאומטריים מהונדסים אלו (על נתונים מחוץ לדגימה) הוא בעל חשיבות עליונה *לפני* הסתמכות מלאה עליהם במסגרת ה-MARL. גישה מדורגת, שבה מערכת ה-MARL הבסיסית מפותחת תחילה ורק לאחר מכן מועשרת בשכבת ניהול הסיכונים הגיאומטרית (לאחר אימותה), עשויה להיות זהירה יותר.

**טבלה 2.1: רכיבי אסטרטגיה יסודיים ומקורות נתונים**

| **שם הרכיב** | **מסגרת זמן (מקור נתונים)** | **אות/מאפיין לאסטרטגיה** | **תפקיד באסטרטגיה/צורה הרמונית** |
| --- | --- | --- | --- |
| MLMI (Machine Learning Momentum Index) | 30 דקות הייקן אשי | 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 | שינוי כיוון עקומת הרגרסיה (שורי/דובי) | אות ראשוני/מאשר בארבע הסינרגיות; קודקוד במשולש ההרמוני |
| FVG (Fair Value Gap) - LuxAlgo | 5 דקות (מותאם להייקן אשי או מחושב על נתונים גולמיים ומיושר) | "מיטיגציה" של FVG (חזרת המחיר למילוי הפער) | אות מאשר בארבע הסינרגיות; קודקוד במשולש ההרמוני |
| LVN (Low Volume Nodes) | פרופיל שוק מנתוני 30 דקות (מותאם להייקן אשי או מחושב על נתונים גולמיים ומיושר) | "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) | מאפיינים גיאומטריים (זוויות, יחסי צלעות, שטח וכו') | משמש להערכת "איכות" הכניסה ולהשפיע על גודל הפוזיציה |

## **3. מערכת משנה לזיהוי מצבי שוק ("MARL קטנה")**

### **3.1. זיהוי מצבי שוק דינמי באמצעות MMD מתקדם**

המערכת תכלול רכיב ייעודי לזיהוי שבעה מצבי שוק מוגדרים, תוך שימוש במדד Higher-Rank Maximum Mean Discrepancy (MMD) הפועל על נתוני Heiken Ashi של 30 דקות.1 גישה זו נבחרה בשל יכולתה ללכוד תלות בנתיב (path-dependency) ודינמיקות לא-מרקוביות המאפיינות סדרות עתיות פיננסיות.1 המימוש המעשי יתבסס על ספריות קוד פתוח כגון issaz/signature-regime-detection ו-higherOrderKME.1

השאילתה המקורית מתייחסת לרכיב זה כאל "מערכת MARL קטנה". עם זאת, התיאורים הנוכחיים במקורות 1 מתמקדים ב-MMD כטכניקה סטטיסטית לזיהוי משטרים, ולא בהכרח כמערכת למידת חיזוק עם סוכנים הלומדים מדיניות לקבלת החלטות סדרתיות. אם הכוונה היא למערכת סיווג מבוססת MMD (עם פרמטרים הניתנים לכיול, אולי באמצעות Optuna כפי שהוצע ב-1), הרי שמדובר במודול אנליטי. אם אכן מדובר במערכת MARL, משמעות הדבר היא שסוכנים לומדים מדיניות לסיווג או לתגובה למצבי שוק, ייתכן במטרה למטב את איכות המידע על מצב השוק המועבר למערכת ה-MARL "הגדולה". בהיעדר פירוט נוסף על מנגנון ה-RL ב"מערכת הקטנה", ההנחה הרווחת כעת היא שמדובר במערכת סיווג מבוססת MMD. יש להבהיר האם שבעת מצבי השוק הם קטגוריות קבועות מראש המבוססות על ספי MMD, או שהם קטגוריות נלמדות. אם הם נלמדים, יש להגדיר את מנגנון הלמידה (למשל, אשכולות (clustering) בלתי מונחה על פלטי MMD).

### **3.2. ממשק אינטגרציה עם מערכת ה-MARL הראשית למסחר**

הפלט של מערכת ה-MMD (תווית המצב מבין השבעה או ציון MMD רציף) ישמש כמאפיין קלט ישיר לייצוג המצב של הסוכנים במערכת ה-MARL "הגדולה".1 מידע זה על מצב השוק הוא חיוני ליכולתם של סוכני ה-MARL "הגדולים" להתאים את המדיניות המשותפת שלהם לתנאי שוק משתנים. תדירות העדכון של מצב ה-MMD (למשל, כל 30 דקות) תקבע את תדירות השינוי של מאפיין הקשר זה עבור הסוכנים הראשיים.1

**טבלה 3.1: 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. |

## **4. מערכת למידת חיזוק מרובת סוכנים (MARL) ראשית ("MARL גדולה")**

### **4.1. התמחות סוכנים: זרמי נתונים קלט והנדסת מאפיינים**

המערכת תכלול סוכנים ייעודיים עבור נתוני 30 דקות, נתוני 5 דקות, אינדיקטורים, ניתוח LVN, וסוכן המקושר למערכת זיהוי מצבי השוק [User Query]. גישה זו מרמזת שכל סוכן מקבל "תצוגה" או תת-קבוצה ייחודית של המידע הכולל הזמין.1

מבנה מטריצת קלט (NxF) לכל סוג סוכן:

השאילתה מציינת סוכנים ייעודיים ל"אינדיקטורים" ו-"LVN". עם זאת, אינדיקטורים (MLMI, NW-RQK, FVG) וניתוח LVN הם גם חלק אינטגרלי מזרמי הנתונים של 30 ו-5 דקות ומהאסטרטגיה המרכזית. אם, לדוגמה, סוכן ה-30 דקות כבר מעבד את נתוני MLMI, NW-RQK ו-LVN מהקלט שלו, תפקידם של סוכני "אינדיקטורים" ו-"LVN" נפרדים הופך פחות ברור ועלול להוביל לכפילות עיבוד או לפרשנויות סותרות.

כדי למנוע כפילות ולהבטיח שכל המידע ההכרחי מעובד ביעילות, מוצע מבנה סוכנים המתמחה לפי *מקור נתונים עיקרי והאלמנטים האסטרטגיים הנגזרים ממנו*:

1. **סוכן מסגרת זמן 30 דקות (ATF30):** קלט: מטריצת N30​×F30​ מבוססת נתוני HA OHLCV של 30 דקות. מעבד/מפיק: אותות MLMI, NW-RQK, ומאפייני LVN.
2. **סוכן מסגרת זמן 5 דקות (ATF5):** קלט: מטריצת N5​×F5​ מבוססת נתוני HA OHLCV של 5 דקות. מעבד/מפיק: אותות FVG.
3. **סוכן מצב שוק (AMS):** קלט: פלט זיהוי המצבים ממערכת ה-MMD (למשל, תווית מצב 1 מתוך 7).
4. **סוכן סיכון הרמוני (AHR):** קלט: אותות הכניסה (זמנים ומחירים) מ-ATF30 ו-ATF5, המשמשים ליצירת "המשולשים ההרמוניים". מחשב את המאפיינים הגיאומטריים של משולשים אלו. תפקידו במסגרת המדיניות המשותפת יהיה ככל הנראה לווסת את גודל הפוזיציה או את רמות ה-SL/TP שנקבעו על סמך פרשנויות של סוכנים אחרים. מבנה זה עולה בקנה אחד עם דרישת המשתמש ל"רכיב חיצוני לניתוח 'צורות הרמוניות'... שישמש את הסוכנים ללימוד ניהול סיכונים דינמי".

**הנדסת מאפיינים למטריצת NxF:** כפי שפורט ב-1, כל מטריצת קלט תכלול נתוני HA, ערכי אינדיקטורים רלוונטיים, מאפייני LVN, גיאומטריית משולש הרמוני (אם רלוונטי לסוכן), ופלט מצב ה-MMD. נרמול וסקיילינג של המאפיינים הם חיוניים.1

**טיפול במאפיינים מבוססי-אירועים/ספורדיים:** ייצוג של משולשים הרמוניים (שאינם נוצרים בכל צעד זמן) יעשה שימוש בדגלים בינאריים, נתוני האירוע האחרון הנישאים קדימה, ומאפייני "זמן מאז האירוע האחרון".1

### **4.2. ארכיטקטורת רשת הטמעה: רשתות משותפות לעומת נפרדות**

מטרת רשת ההטמעה היא להמיר את מטריצת הקלט Mi​ של כל סוכן (שעשויה להיות הטרוגנית ובעלת ממדיות גבוהה) לייצוג וקטורי דחוס (embedding) בעל ממדיות נמוכה יותר, אשר ישמש קלט לרשת המדיניות.1 ארכיטקטורות אפשריות כוללות 1D-CNNs, LSTMs/GRUs, Transformers, ואוטואנקודרים (AE/VAE), או שילובים היברידיים שלהם.1

**השאלה המרכזית היא האם להשתמש ברשת הטמעה אחת משותפת לכל הסוכנים, או ברשתות נפרדות ומותאמות לכל סוג סוכן.**

* **רשת הטמעה משותפת:**
  + *יתרונות:* עשויה לקדם "שפה משותפת" או מרחב מאפיינים עקבי בין הסוכנים, מה שעשוי לשפר תיאום אם סוכנים צריכים להבין באופן מרומז את מצביהם של אחרים.1 יכולה להיות יעילה יותר מבחינת מספר הפרמטרים אם המשקולות משותפות.
  + *חסרונות:* עלולה שלא להיות אופטימלית אם מטריצות הקלט Mi​ של הסוכנים השונים הן בעלות מבנים, סקלות או סוגי מידע שונים מאוד.1 ארכיטקטורה יחידה עלולה להתקשות בעיבוד יעיל של כל סוגי הקלט המגוונים.
* **רשתות הטמעה נפרדות:**
  + *יתרונות:* מאפשרות עיבוד אופטימלי של זרם הנתונים הייחודי לכל סוכן. לדוגמה, סוכן המעבד נתוני מחיר גולמיים ב-5 דקות עשוי להפיק תועלת מ-CNN, בעוד שסוכן המעבד ערכי אינדיקטורים שכבר חושבו עשוי להשתמש ב-MLP או LSTM פשוט יותר.1
  + *חסרונות:* מורכבות יותר לתכנון ולתחזוקה. פחות שיתוף ישיר של ייצוגים נלמדים בין הסוכנים.

המלצה על ארכיטקטורת רשת ההטמעה:

בהתחשב באופי הייחודי של הקלט לכל סוג סוכן שהוצע (סדרות עתיות גולמיות עבור ATF30/ATF5, תווית מצב קטגורית עבור AMS, מאפיינים גיאומטריים עבור AHR), ארכיטקטורות רשת הטמעה נפרדות, המותאמות לכל סוג סוכן, הן ככל הנראה המתאימות ביותר.

* סוכני ATF30 ו-ATF5 (המעבדים מטריצות סדרתיות N×F) יכולים להשתמש במקודדים מבוססי Transformer או LSTM.
* סוכן AMS (המקבל תווית מצב בודדת או ציון) עשוי לדרוש שכבת הטמעה פשוטה יותר (למשל, טבלת חיפוש הטמעה אם הקלט קטגורי, או העברה ישירה/MLP אם הקלט הוא ציון).
* סוכן AHR (המקבל מאפיינים גיאומטריים) עשוי להשתמש ב-MLP. עם זאת, אם הקלט המבני של ATF30 ו-ATF5 דומה מאוד (למרות מסגרות הזמן השונות), הם *יכולים* לחלוק *תבנית ארכיטקטונית* (למשל, מקודד Transformer) אך עם משקולות הנלמדות בנפרד כדי לאפשר התמחות.

אימון רשת ההטמעה יתבצע מקצה-לקצה עם פונקציית ההפסד של סוכן ה-RL (למשל, שגיאת TD), כפי שהועדף על ידי המשתמש 1, מכיוון שהדבר הופך את ההטמעה לספציפית למשימה. אימון מקדים (למשל, באמצעות אוטואנקודר) יכול לשמש כשלב אתחול.1

עבור סוכנים המעבדים נתונים ממסגרות זמן שונות (ATF30 ו-ATF5), כאשר מטריצת הקלט N×F מבוססת על רזולוציית 5 דקות ומכילה מאפייני הקשר של 30 דקות (הקבועים למשך שישה צעדי 5 דקות 1), יש לשקול ארכיטקטורות הטמעה היררכיות. ארכיטקטורה כזו יכולה לכלול רכיב אחד המתמקד במאפייני 5 הדקות המשתנים במהירות, ורכיב אחר המעבד את מאפייני 30 הדקות המשתנים לאט יותר, כאשר הפלטים שלהם משולבים לאחר מכן. גישה זו מאפשרת לרשת ללמוד ייצוגים בסקלות זמניות מתאימות. ארכיטקטורות Transformer עם קידוד מיקום (positional encoding) מתאים עשויות להתמודד עם מבנה היררכי זה באופן טבעי.

### **4.3. אימון מרכזי עם ביצוע מבוזר (CTDE) באמצעות MAPPO**

* **פרדיגמה:** CTDE מתאימה היטב ל-MARL שיתופי, שבו סוכנים חולקים מטרה משותפת אך פועלים על סמך תצפיות חלקיות.1 המבקר (critic) הוא מרכזי במהלך האימון (בעל גישה למצב גלובלי או לתצפיות/פעולות משולבות של הסוכנים), בעוד שהשחקנים (actors - המדיניויות) מבוזרים במהלך הביצוע.
* **MAPPO (Multi-Agent PPO):** אלגוריתם שהוגדר על ידי המשתמש [User Query]. הוא מרחיב את PPO ל-MARL ומתאים למדיניות משותפת ולמבקר מרכזי.1 ידוע ביציבותו.
* **מדיניות משותפת:** כל הסוכנים המתמחים פועלים תחת רשת מדיניות משותפת.1 המדיניות ממפה את התצפית המקומית (המוטמעת) si​ של הסוכן לפעולה ai​.
  + **שיתוף פרמטרים:** שיתוף פרמטרים ברשת המדיניות בין הסוכנים נפוץ ב-MAPPO, במיוחד עבור סוכנים הומוגניים, מכיוון שהוא משפר את יעילות הדגימה.1 בהתחשב בהתמחות הסוכנים כאן, שיתוף פרמטרים מלא עשוי להיות מגביל מדי. שיתוף חלקי (למשל, שכבות תחתונות משותפות, שכבות עליונות מתמחות) או שימוש באותה ארכיטקטורה עם משקולות שונות יכולים להוות פשרה. מחקר 16 דן במיפוי סוגי סוכנים לאזורים שונים ברשת משותפת.
* **מבקר מרכזי:** המבקר ב-MAPPO מקבל בדרך כלל את המצב הגלובלי (או שרשור של כל תצפיות הסוכנים ו/או פעולותיהם) כקלט כדי לייצר הערכת ערך יציבה יותר, ובכך למתן את אי-הנייחות הנובעת מלמידה של סוכנים אחרים.1

### **4.4. ניסוח מדיניות משותפת ועיצוב מרחב פעולות**

* **רשת מדיניות:** סביר שתהיה MLP הפועל על וקטור המצב המוטמע מכל סוכן.
* **מרחב פעולות (היברידי):**
  + בדיד: כיוון עסקה (לונג, שורט, החזק/ללא פעולה).1
  + בדיד/רציף: גודל פוזיציה (מושפע מחוזק LVN ואיכות הרמונית).1
  + רציף: רמות Stop-Loss (SL) ו-Take-Profit (TP) (דינמיות, מושפעות מניתוח הרמוני) [User Query].
* **פרשנות "ארבע הסינרגיות":** המערכת אינה בוחרת סינרגיה. השוק מכתיב את רצף אותות MLMI, NW-RQK, FVG. המדיניות לומדת להגיב באופן אופטימלי לאחר שצירוף אותות תקף (אחת ה"סינרגיות") התממש ויצר משולש הרמוני.1 הרצף/הגיאומטריה הספציפיים הופכים לחלק ממצב הסוכן.

### **4.5. היתכנות ומנגנונים לשיתוף מידע בין סוכנים בזמן אמת**

השאילתה המרכזית כאן היא היתכנותו של שיתוף מידע בזמן אמת בין הסוכנים.

עקרון ה-CTDE גורס כי הביצוע המבוזר לרוב אינו כולל תקשורת ישירה בזמן אמת בין השחקנים, כדי לשמור על פשטות ולהימנע מצווארי בקבוק בתקשורת.1 התיאום נלמד באופן מרומז דרך המבקר המרכזי והמדיניות המשותפת.

* **טיעונים בעד שיתוף (שיפור תיאום):** בשווקים פיננסיים מורכבים, שיתוף תצפיות חלקיות או כוונות פעולה עשוי לשפר את שיתוף הפעולה ולהפחית החלטות סותרות, במיוחד אם התצוגות המקומיות אינן מספקות דיין.17
* **טיעונים נגד שיתוף (הפרת עקרון הביצוע המבוזר, מורכבות):** מגדיל את מורכבות המערכת, את ההשהיה הפוטנציאלית ואת נקודות הכשל. מקשה על בעיית הלמידה, שכן הקלט של סוכנים הופך תלוי בהודעות (שעשויות להיות רועשות או מושהות) של סוכנים אחרים. MAPPO עם מבקר מרכזי כבר מספק מנגנון רב עוצמה לתיאום מרומז. כפי שצוין ב-1, אם מידע קריטי למספר סוכנים, הוא אמור להיות חלק מהתצפית המקומית שלהם או מטופל על ידי שכבות הטמעה משותפות, ולא באמצעות העברת הודעות פעילה.
* **חלופה: שיתוף מרומז דרך רכיבי מצב גלובליים:**
  + מצב השוק מ-MMD הוא כבר צורה של מידע גלובלי משותף.
  + מאפייני "המשולש ההרמוני", אם מחושבים באופן מרכזי ומוזנים לסוכנים הרלוונטיים, פועלים גם הם כמידע הקשרי משותף הנגזר ממספר מקורות אותות.

רכיב "ניתוח הצורות ההרמוניות" (שהוצע כסוכן AHR) משלב מטבעו אותות ממקורות ראשוניים שונים (MLMI מ-ATF30, FVG מ-ATF5 וכו') ליצירת משולשים. המאפיינים הגיאומטריים הנגזרים על ידי AHR משמשים לאחר מכן את הסוכנים לניהול סיכונים. באופן זה, AHR פועל כנקודת מיזוג ותיאום מידע מרומזת. הוא לוקח מידע שעשוי להגיע מסוכנים "מתמחים" שונים (או מזרמי הנתונים שלהם) ומסנתז אותו לתובנה הקשרית חדשה ומשותפת (הגיאומטריה והאיכות של המשולש). עיצוב זה מיישם באופן טבעי צורה של שיתוף מידע מבלי לדרוש העברת הודעות ישירה בין סוכני הנתונים הראשיים (ATF30, ATF5). "המידע המשותף" הוא התובנה הגיאומטרית המזוקקת המשמשת לניהול סיכונים.

המלצה:

בשלב ראשון, יש להימנע מהעברת הודעות ישירה בזמן אמת בין שחקני הסוכנים. יש להתמקד בתיאום מרומז חזק באמצעות מסגרת CTDE/MAPPO, מדיניות משותפת, וייצוגי מצב מקיפים הכוללים הקשר גלובלי (מצב MMD, מאפיינים הרמוניים). אם הביצועים ידגימו צורך ברור בתיאום מפורש יותר, ניתן יהיה לבחון:

* העברת מידע מוגבלת ומובנית: למשל, סוכן AHR המשדר פרמטרי סיכון גיאומטריים מרכזיים.
* מנגנוני קשב במבקר או במדיניות: המאפשרים לסוכנים לשקלל מידע מסוכנים (רעיוניים) אחרים במהלך קבלת החלטות, כפי שהוצע ב-18, אם מנגנון ה"סופרפוזיציה" יתפתח לכיוון זה.

**טבלה 4.1: התמחות סוכני MARL ומבנה קלט/פלט**

| **שם הסוכן** | **קלט עיקרי** | **אחריות עיקרית / מאפיינים נגזרים** | **תרומת פלט למדיניות/פעולה משותפת** |
| --- | --- | --- | --- |
| ATF30 (סוכן מסגרת זמן 30 דקות) | מטריצת N30​×F30​ מנתוני HA של 30 דקות | הפקת אותות MLMI, NW-RQK; זיהוי LVN והערכת "חוזק תחנה" | מספק מידע על מגמה ארוכת-טווח ורמות מבניות |
| ATF5 (סוכן מסגרת זמן 5 דקות) | מטריצת N5​×F5​ מנתוני HA של 5 דקות | הפקת אותות FVG | מספק מידע על חוסר איזון קצר-טווח ותזמון כניסה פוטנציאלי |
| AMS (סוכן מצב שוק) | פלט זיהוי מצבים ממערכת MMD (תווית/ציון) | העברת הקשר שוק גלובלי | מאפשר למדיניות המשותפת להתאים עצמה למצב השוק |
| AHR (סוכן סיכון הרמוני) | אותות כניסה מ-ATF30 ו-ATF5 (זמנים ומחירים) | יצירת "משולשים הרמוניים", חישוב מאפיינים גיאומטריים | מספק הערכת "איכות" של מערך הכניסה, משפיע על פרמטרי ניהול סיכונים (גודל פוזיציה, SL/TP) |

**טבלה 4.2: ניתוח השוואתי של רשתות הטמעה משותפות לעומת נפרדות** (מבוסס על 1)

| **גישה** | **יתרונות** | **חסרונות** | **השלכות על יציבות אימון** | **פוטנציאל ביצועים** | **מדרגיות** | **התאמה לקלט הטרוגני** |
| --- | --- | --- | --- | --- | --- | --- |
| רשת משותפת - משקולות משותפות לחלוטין | יעילות פרמטרים גבוהה, מקדמת "שפה משותפת" חזקה. | עלולה להיות מגבילה מדי עבור קלטים הטרוגניים מאוד, קושי בהתמחות. | עשויה להיות יציבה יותר אם הקלטים דומים, אך לא יציבה אם נאלצת להתמודד עם סוגי מידע שונים מדי. | מוגבלת על ידי יכולת ההכללה של רשת יחידה על פני כל סוגי הקלט. | גבוהה מבחינת מספר פרמטרים. | נמוכה. |
| רשת משותפת - ארכיטקטורה משותפת, משקולות נפרדות | מאפשרת התמחות מסוימת תוך שמירה על מבנה אחיד, עדיין מקדמת עקביות מסוימת. | פחות יעילה מבחינת פרמטרים מרשת משותפת לחלוטין. | יציבות תלויה ביכולת הארכיטקטורה להתאים למגוון קלטים. | טובה יותר מרשת משותפת לחלוטין עבור קלטים מגוונים במידה. | בינונית. | בינונית. |
| רשתות נפרדות לכל סוג סוכן | מאפשרת אופטימיזציה מלאה של ההטמעה לכל סוג קלט ספציפי, גמישות ארכיטקטונית מקסימלית. | יותר פרמטרים כוללים, פחות שיתוף ישיר של ייצוגים נלמדים, מורכבות תכנונית גבוהה יותר. | יציבות אימון תלויה בכל רשת בנפרד; עשויה להיות קלה יותר לכוונון עבור כל סוכן. | פוטנציאל ביצועים גבוה ביותר אם כל רשת מותאמת היטב. | נמוכה יותר מבחינת יעילות פרמטרים, אך מדרגית מבחינת הוספת סוגי סוכנים חדשים. | גבוהה. |

## **5. אגרגציית החלטות היררכית: מנגנון ה"סופרפוזיציה"**

### **5.1. בסיס רעיוני: מהשראה קוונטית למימוש קלאסי**

המשתמש הציג רעיון של "סופרפוזיציה" בהשראת מכניקת הקוונטים, במטרה לשפר את עמידות ואמינות המערכת.1 חשוב להדגיש כי אין הכוונה למימוש חישוב קוונטי אמיתי, אלא לשאוב השראה רעיונית ולממשה באמצעות אלגוריתמים קלאסיים.1 אנלוגיות קלאסיות אפשריות כוללות שימוש באנסמבל של מדיניויות, מדיניות הסתברותית המפיקה התפלגות על פני פעולות, או מנגנון דמוי-QIDDM שבו פעולות/סינרגיות פוטנציאליות מוערכות במקביל, והמערכת "קורסת" להחלטה ספציפית עם קבלת טריגר מתאים.1

### **5.2. ארכיטקטורה מבוססת קשב (Attention) למיזוג החלטות סוכנים**

הדרישה היא למנגנון "סופרפוזיציה" מבוסס קשב לאגרגציית החלטות הסוכנים [User Query]. במקום הצבעה פשוטה או ממוצע, מנגנון קשב יכול לשקלל באופן דינמי את ה"דעות" או הפלטים של הסוכנים המתמחים השונים (ATF30, ATF5, AMS, AHR) בעת גיבוש החלטת מסחר סופית.

* כל סוכן i מייצר ייצוג פנימי או המלצת פעולה/ערך ראשונית hi​, המבוססת על התצוגה הייחודית שלו Mi​ והמדיניות המשותפת.
* מודול קשב מקבל ייצוגים אלו {h1​,h2​,...,hk​} מ-k הסוכנים כקלט.
* המודול לומד להקצות משקולות קשב αi​ לכל hi​, המשקפות את הרלוונטיות או הביטחון שלו בהקשר הנוכחי.
* ההחלטה או הייצוג הסופי המאוגרד הוא סכום משוקלל: hfinal​=∑(αi​⋅hi​).
* hfinal​ זה ימופה לאחר מכן לפעולת מסחר קונקרטית.

מנגנוני קשב ב-MARL הוצעו בהקשרים של תקשורת בין-סוכנית או פירוק פונקציות ערך.18 אף שאינם מיועדים ישירות לאגרגציית החלטות בתצורת CTDE זו, העיקרון של שימוש בקשב לשקלול תרומות ממקורות שונים הוא רלוונטי ביותר. שכבת האגרגציה המרכזית שהוזכרה ב-1 תתממש באמצעות מנגנון זה.

ניתן לאמן את מנגנון אגרגציית הקשב מקצה-לקצה כחלק ממערכת ה-MARL הכוללת. התגמולים מפעולות המסחר יתפשטו לאחור דרך המאגרגד ואל הסוכנים הבודדים ורשתות ההטמעה שלהם.

יש לשים לב כי אם כל הסוכנים משתמשים באותה רשת מדיניות והיא דטרמיניסטית בהינתן מצבם (המוטמע), הם יפיקו פעולות זהות אם מצביהם זהים. ההבדלים בפעולות נובעים מהבדלים בתצוגות הקלט שלהם (Mi​), המובילים למצבים מוטמעים שונים. לפיכך, אם מנגנון ה"סופרפוזיציה"/קשב נועד למזג *החלטות* או *פרספקטיבות* שונות, הבדלים אלו חייבים לנבוע מהקלטים המתמחים של הסוכנים, ולא מכך שהם מריצים מדיניויות שונות במהותן (אלא אם ה"סופרפוזיציה" כוללת הרצת אנסמבל של גרסאות של המדיניות המשותפת, כפי שהוצע ב-1). במקרה הראשון, מנגנון הקשב ילמד לשקלל את חשיבות המידע הנגזר מזרמי נתונים שונים (30 דקות, 5 דקות, מצב MMD, מאפייני סיכון הרמוניים), כפי שעובדו על ידי המדיניות המשותפת דרך "עדשות" של סוכנים שונים. זהו מיזוג מתוחכם של ראיות, ולא אגרגציה של "הצבעות סותרות מאיי-איי שונים".

### **5.3. הבטחת פעולות מסחר מערכתיות קוהרנטיות**

הפלט של מאגרגד הקשב חייב להיות החלטת מסחר יחידה וניתנת לביצוע (למשל, קנה X חוזים, הצב SL ב-Y, TP ב-Z). התכנון חייב למנוע מצבים שבהם אותות מאוגרדים הם דו-משמעיים או מובילים לפעולות סותרות ברמה נמוכה. הדבר מטופל בדרך כלל על ידי השכבה הסופית של רשת המאגרגד, הממפה לייצוג הפעולה המוגדר. פונקציית התגמול הכוללת (סעיף 6.2) תהיה קריטית בעיצוב המאגרגד להפקת פעולות קוהרנטיות ורווחיות.

## **6. אימון מערכת, אימות ואופטימיזציה**

### **6.1. הכנת נתונים ומסגרת אופטימיזציה Walk-Forward (WFO)**

* **חלוקת נתונים:** חמש שנות נתונים היסטוריים. חלוקה כרונולוגית חיונית לסדרות עתיות.1
* **WFO:** נחשבת ל"תקן הזהב" לאימות אסטרטגיות.21 כוללת אופטימיזציה/אימון על חלון נתונים בתוך-הדגימה (In-Sample - IS), ולאחר מכן בדיקה על חלון נתונים מחוץ-לדגימה (Out-of-Sample - OOS) עוקב מיידית, והזזת תהליך זה קדימה.1
* **WFO עם MARL:** כל מקטע WFO יכלול:
  + אימון מערכת ה-MMD (כיול פרמטרים או למידת הגדרת מצבים) על נתוני ה-IS.
  + אימון מערכת ה-MARL "הגדולה" (רשתות הטמעה, מדיניות משותפת, מאגרגד קשב) על נתוני ה-IS, תוך שימוש במצבי ה-MMD שנוצרו על אותם נתוני IS.
  + הערכת המערכת כולה על נתוני ה-OOS.

אימון מערכת MARL מתוחכמת הוא עתיר חישוב. ביצוע אימון MARL מלא בתוך כל חלון IS של תהליך WFO, במיוחד אם נעשה שימוש במקטעים רבים, עלול להוביל לזמני אופטימיזציה ארוכים במיוחד. יש צורך באסטרטגיות לניהול עלות זו, כגון שימוש באלגוריתמי MARL יעילים (MAPPO נחשב יחסית טוב), יכולות אימון מבוזר (למשל, Ray RLlib), תדירות נמוכה יותר של אימון-מחדש מלא ב-WFO (עם עדכונים או כוונון עדין תכופים יותר של רכיבים מסוימים), ובחירה קפדנית של גודלי חלונות ה-WFO וצעדי ההזזה. היתכנות מעשית של WFO קפדני תלויה במשאבים החישוביים הזמינים וביעילות לולאת האימון של ה-MARL.

### **6.2. הנדסת פונקציית תגמול ל-MARL שיתופי**

* **תגמול צוות גלובלי:** כל הסוכנים במערכת ה-MARL "הגדולה" חולקים אות תגמול משותף המשקף את ביצועי התיק הכוללים.1
* **רכיבים** 1:
  + רווח והפסד (P&L) של התיק (עיקרי).
  + תשואה מותאמת-סיכון (יחס שארפ/סורטינו).
  + עונשים על משיכה מקסימלית (drawdown) מופרזת, תנודתיות גבוהה.
  + הכללת עלויות עסקה (עמלות, החלקה).
  + (אופציונלי) בונוס על ביצוע מוצלח של "סינרגיות" או עמידה בפרמטרי סיכון רצויים מניתוח הרמוני (אם כי ייתכן שעדיף שזה יילמד באופן מרומז אם הדבר מוביל ל-P&L/שארפ טובים יותר).
* **הקצאת אשראי:** בעוד שנעשה שימוש בתגמול גלובלי, המבקר המרכזי ב-MAPPO מסייע בהקצאת אשראי/אשמה במהלך האימון על ידי הערכת פעולות בהקשר של המצב הגלובלי.1 ניתן לבחון בהמשך שיטות פירוק ערך מפורשות (VDN, QMIX) אם יתעורר צורך.1

### **6.3. אופטימיזציית היפר-פרמטרים באמצעות Optuna**

* **כלים:** Optuna, עם אפשרות לשילוב עם Ray Tune לאופטימיזציית היפר-פרמטרים מבוזרת.1
* **היקף:** אופטימיזציה של היפר-פרמטרים עבור:
  + אלגוריתם ה-MARL (שיעורי למידה, גודלי אצווה, ארכיטקטורות רשת למדיניות/מבקר/הטמעה).
  + מודול זיהוי מצבי MMD (פרמטרי ליבה, גודלי חלונות, ספים).1
  + מאגרגד "סופרפוזיציה" מבוסס קשב.
  + אורך הרצף N עבור מטריצות הקלט.
* **פונקציית מטרה ל-HPO:** בדרך כלל מדד ביצועים חזק משלב ה-OOS של WFO (למשל, ממוצע יחס שארפ על פני מקטעי WFO).

קיימת תלות הדדית בין אופטימיזציית היפר-פרמטרים (HPO) לבין WFO. ההיפר-פרמטרים "הטובים ביותר" שיימצאו על ידי Optuna עשויים להיות רגישים בעצמם לנתוני ה-IS הספציפיים של מקטע WFO מסוים. הרצת HPO *בתוך כל מקטע IS* של ה-WFO היא הגישה הקפדנית ביותר אך יקרה מאוד חישובית. גישה מעשית יותר עשויה לכלול הרצת HPO על מספר מקטעי WFO ראשוניים כדי למצוא קבוצת היפר-פרמטרים טובה באופן כללי, ואז להשתמש בהיפר-פרמטרים קבועים אלו עבור שאר מקטעי ה-WFO כדי להעריך רובסטיות. הדבר מדגיש כי פרמטרים "אופטימליים" עשויים שלא להיות סטטיים; באופן אידיאלי, המערכת צריכה להיות מתוכננת כך שתהיה רגישה יחסית פחות לשינויים קלים בהיפר-פרמטרים, או ש-HPO יהיה חלק מצנרת למידה מתמשכת.

**טבלה 6.1: פרוטוקול אופטימיזציה Walk-Forward ומדדי מפתח** (בהשראת 1)

| **איטרציית WFO** | **תקופת In-Sample (IS)** | **תקופת Out-of-Sample (OOS)** | **משימות אימון/אופטימיזציה ב-IS** | **מדדי הערכה מרכזיים ב-OOS** |
| --- | --- | --- | --- | --- |
| 1 | שנים 1-2 | שנה 2.25-2.5 | אימון ראשוני של MMD, MARL (הטמעות, מדיניות, אגרגטור), HPO (אם מתבצע פר-מקטע) | יחס שארפ, יחס סורטינו, MDD, P&L, שיעור זכייה, שיעור ביצוע סינרגיות |
| 2 | שנים 1.25-2.25 | שנה 2.5-2.75 | אימון מחדש/כוונון עדין של MMD, MARL. עדכון HPO (אם רלוונטי) | יחס שארפ, יחס סורטינו, MDD, P&L, שיעור זכייה, שיעור ביצוע סינרגיות |
| ... | ... | ... | ... | ... |
| N | שנים X-(Y+2) | שנים (Y+2)-(Y+2.25) | אימון מחדש/כוונון עדין אחרון. | יחס שארפ, יחס סורטינו, MDD, P&L, שיעור זכייה, שיעור ביצוע סינרגיות |

## **7. ארכיטקטורת פריסה ושיקולים תפעוליים למסחר חי**

### **7.1. תשתית: Windows VPS לעומת פתרונות מבוססי ענן**

השאילתה המרכזית היא בחירה בין Windows VPS לבין פתרונות ענן עבור מסחר בחוזים עתידיים.

* **Windows VPS (Virtual Private Server):**
  + *יתרונות:* לעיתים קרובות מועדף בשל השהיה נמוכה יותר אם ה-VPS ממוקם בקרבה פיזית (co-location) או קרוב מאוד לשרתי הבורסה.23 יכול להציע משאבים ייעודיים, גישת root מלאה, וחלק מהספקים מבצעים אופטימיזציה מראש עבור פלטפורמות מסחר (למשל, NinjaTrader, MetaTrader ב-QuantVPS 24).
  + *חסרונות:* מדרגיות יכולה להיות פחות גמישה מאשר בענן. המשתמש אחראי יותר לתחזוקת מערכת ההפעלה והתוכנה. האמינות תלויה מאוד בהתחייבויות הספק לזמן פעולה תקינה (QuantVPS טוען ל-100% 24).
* **פתרונות ענן (AWS, Azure, GCP):**
  + *יתרונות:* מדרגיות וגמישות גבוהות. מגוון רחב של שירותים (מסדי נתונים, סביבות הרצת ML). פוטנציאל לטווח הגעה גלובלי טוב יותר אם סוחרים במספר בורסות. שירותים מנוהלים יכולים להפחית את תקורת התחזוקה.
  + *חסרונות:* ההשהיה לבורסות עלולה להיות גבוהה יותר ומשתנה יותר אם לא משתמשים בשירותי חיבור ישיר (direct connect) או co-location שמציעים ספקי הענן במרכזי נתונים פיננסיים. עלויות יכולות להיות קשות יותר לחיזוי ויכולות לעלות עם השימוש.

עבור מסחר בנכס יחיד בחוזים עתידיים, כאשר השהיה נמוכה לבורסה ספציפית (למשל, CME) היא בעלת חשיבות עליונה, **Windows VPS איכותי הממוקם ב-co-location עם מנוע ההתאמה של הבורסה או בקרבה פיזית רבה אליו הוא בדרך כלל הבחירה המועדפת.** QuantVPS ו-FXVM הן דוגמאות לספקים המתמחים בכך.23 אם האסטרטגיה כוללת נכסים מרובים על פני בורסות גלובליות מגוונות או דורשת משאבים חישוביים משמעותיים ואלסטיים למשימות שמעבר להסקה, ניתן לשקול גישת ענן היברידית או ענן עם קישוריות ישירה לבורסה.

השהיה נמוכה היא יתרון מרכזי המוזכר עבור VPS במסחר.23 במסחר בחוזים עתידיים, ההשהיה אינה נוגעת רק למהירות עיבוד השרת אלא בעיקר למרחק הרשת ולאיכות הנתיב בין אלגוריתם המסחר למנוע ההתאמה של הבורסה. לכן, בחירה פשוטה בין "ענן" ל-"VPS" אינה מספקת. *המיקום הפיזי הספציפי* של השרת (VPS או מופע ענן) ביחס לבורסה הוא הגורם הדומיננטי להשהיית ביצוע. ספקי ענן אכן מציעים מופעים במרכזי נתונים פיננסיים (למשל, AWS בשיקגו עבור CME), אך יש לבחור אותם באופן ספציפי. ההחלטה כרוכה בפשרה: ספקי VPS ייעודיים למסחר לעיתים קרובות מבצעים אופטימיזציה עבור קרבה זו, בעוד שספקי ענן כלליים מציעים גמישות רחבה יותר אך דורשים תצורה קפדנית ליישומי פיננסים עם השהיה נמוכה. טענת "השהיה של 0.7ms" על ידי QuantVPS 24 מתייחסת ספציפית לקרבה לשיקגו/CME.

### **7.2. הסקה בביצועים גבוהים: מינוף ONNX ו-TensorRT**

הדרישה היא להשתמש ב-ONNX/TensorRT להסקה מהירה.

* **ONNX (Open Neural Network Exchange):** פורמט פתוח לייצוג מודלי ML. ניתן לייצא מודלי PyTorch (הטמעה, מדיניות, אגרגטור) ל-ONNX.25
* **TensorRT:** ערכת פיתוח תוכנה (SDK) של NVIDIA להסקת למידה עמוקה בביצועים גבוהים. מבצעת אופטימיזציה של מודלי ONNX עבור מעבדים גרפיים (GPUs) של NVIDIA, מפחיתה משמעותית השהיה ומגדילה תפוקה.25
* **תהליך עבודה:** מודל PyTorch → מודל ONNX → מנוע TensorRT.

שימוש זה חיוני למזעור השהיית קבלת ההחלטות במסחר חי, ומבטיח שהמערכת יכולה להגיב במהירות לשינויי שוק. יש לקחת בחשבון שלא כל פעולות PyTorch ניתנות לייצוא ישיר ל-ONNX או לאופטימיזציה על ידי TensorRT, במיוחד פעולות מותאמות אישית או חדשות מאוד. ייתכן שיידרש פישוט או ניסוח מחדש של המודל.27 אופטימיזציית TensorRT היא ספציפית ל-GPU, ויש לנהל צורות קלט קבועות לעומת דינמיות.27

### **7.3. שילוב Human-in-the-Loop (HITL) לפיקוח ובקרה**

הדרישה היא לכלול HITL. עבור מערכות אוטונומיות מורכבות בסביבות בעלות סיכון גבוה כמו מסחר, HITL חיוני לבטיחות, אמון, וטיפול במצבים בלתי צפויים.1

* **תפקידי HITL** 1:
  + ניטור התנהגות וביצועי המערכת.
  + אימות זיהויי מצבי MMD או אותות קריטיים.
  + עקיפת החלטות מערכת במצבים חריגים (למשל, "מתג השבתה", הקטנת גודל פוזיציה).
  + הנחיית אימון מחדש או כוונון עדין של מודלים על סמך ביצועים נצפים.
* **ממשק:** נדרש לוח מחוונים (dashboard) לניטור משתני מערכת מרכזיים, מדדי ביצועים, ופוזיציות נוכחיות, עם בקרות להתערבות.

HITL נתפס לעיתים קרובות כמנגנון בטיחות/עקיפה. עם זאת, התערבויות אנושיות, אם הן מתועדות ומנותחות, יכולות לספק נתונים יקרי ערך על מצבים שבהם הבינה המלאכותית נכשלה או התנהגה באופן תת-אופטימלי. מערכת ה-HITL צריכה להיות מתוכננת לא רק להתערבות, אלא גם ללכידת *הסיבות* להתערבות. משוב זה יכול לשמש לאחר מכן לזיהוי חולשות במודלי הבינה המלאכותית או באסטרטגיה, ליצירת נתוני אימון חדשים (למשל, תיוג מצבים שבהם נדרשה עקיפה אנושית), ולהנחות איטרציות עתידיות של פיתוח מודלים ועיצוב פונקציית התגמול. באופן זה, HITL מתפתח ממנגנון בקרה פשוט לחלק אינטגרלי ממחזור שיפור מתמשך, מה שהופך את המערכת הכוללת לעמידה יותר ומותאמת יותר לאינטואיציה של מומחים לאורך זמן. הדבר חיוני לקיימות ארוכת-טווח בשווקים דינמיים.

**טבלה 7.1: השוואת תשתיות פריסה: Windows VPS לעומת ענן**

| **קריטריון** | **Windows VPS (ב-Co-location)** | **ענן גנרי (למשל, AWS/Azure באזורים סטנדרטיים)** | **ענן (מופע באזור פיננסי)** |
| --- | --- | --- | --- |
| השהיה לבורסה | נמוכה מאוד (תת-מילישנייה אפשרית) | גבוהה יותר, משתנה | נמוכה, תלויה בקרבה ובקישוריות |
| מדרגיות | מוגבלת יחסית, דורשת שדרוג ידני | גבוהה מאוד, אלסטית | גבוהה, אלסטית |
| מבנה עלויות | לרוב קבוע חודשית | מבוסס שימוש, יכול להשתנות | יקר יותר ממופע ענן גנרי |
| תקורת תחזוקה | גבוהה יותר (OS, אבטחה) | נמוכה יותר (שירותים מנוהלים) | נמוכה יותר |
| שליטה/התאמה אישית | גבוהה מאוד (גישת root) | בינונית עד גבוהה | בינונית עד גבוהה |
| אמינות/זמן פעולה תקינה | תלוי בספק (יש התחייבויות ל-99.9% עד 100%) | גבוהה מאוד (עם יתירות) | גבוהה מאוד |
| התאמה למסחר בחוזים עתידיים | גבוהה מאוד עבור השהיה נמוכה לבורסה ספציפית | פחות אידיאלי להשהיה נמוכה ללא תצורה מיוחדת | טובה, במיוחד אם נדרשת גמישות ענן |

## **8. המלצות ארכיטקטוניות מקיפות וכיווני התפתחות עתידיים**

### **8.1. תוכנית אב ארכיטקטונית מסונתזת**

הארכיטקטורה הכוללת של מערכת המסחר האלגוריתמית מרובת הסוכנים מורכבת ממספר רכיבים מרכזיים הפועלים בתיאום:

1. **קליטת נתונים ועיבוד מקדים:** קליטת נתוני OHLCV גולמיים של חוזים עתידיים (5 דקות ו-30 דקות), המרתם לנרות Heiken Ashi, וניקויים.
2. **הנדסת מאפיינים:**
   * חישוב אינדיקטורי ליבה: MLMI ו-NW-RQK על נתוני 30 דקות HA; FVG על נתוני 5 דקות HA (או גולמיים, עם יישור).
   * ניתוח פרופיל שוק: זיהוי LVNs מנתוני 30 דקות וכימות "חוזק תחנה".
   * זיהוי מצב שוק: מערכת ה-MMD ("MARL קטנה") מפיקה תווית/ציון מצב שוק (1 מתוך 7) מנתוני 30 דקות HA.
3. **התמחות סוכנים וקלט:**
   * סוכנים ייעודיים (ATF30, ATF5, AMS, AHR) מקבלים מטריצות קלט Mi​ ספציפיות המכילות את המאפיינים הרלוונטיים להם.
   * סוכן AHR מזהה התיישרות אותות (סינרגיות), יוצר "משולשים הרמוניים" מנתוני 5 דקות, ומפיק מאפיינים גיאומטריים.
4. **הטמעה ועיבוד על ידי סוכנים:**
   * כל סוכן מעביר את מטריצת הקלט Mi​ שלו דרך רשת הטמעה (embedding) ייעודית (ארכיטקטורה נפרדת לכל סוג סוכן, נלמדת מקצה-לקצה) לקבלת ייצוג מצב דחוס si​.
5. **מדיניות משותפת (MAPPO/CTDE):**
   * כל הסוכנים פועלים תחת רשת מדיניות משותפת π(ai​∣si​), הממפה את מצבם si​ לפעולה ראשונית או להמלצה.
   * מאפייני הסיכון ההרמוניים (מ-AHR) משפיעים על פרמטרים בפעולה (גודל פוזיציה, SL/TP).
6. **אגרגציית החלטות ("סופרפוזיציה"):**
   * מנגנון מבוסס קשב (Attention) מאחד את הפלטים/המלצות הפעולה מהסוכנים השונים להחלטת מסחר מערכתית אחת.
7. **ביצוע פעולה:** הפעולה המאוחדת מבוצעת בשוק (בסימולציה או במסחר חי).
8. **לולאת משוב:** הסביבה מחזירה תגמול (מבוסס P&L, סיכון וכו') ומצב חדש, והתהליך חוזר על עצמו, תוך עדכון רשתות ההטמעה, המדיניות והאגרגטור.

יש לשים דגש על מודולריות ועל ממשקים מוגדרים היטב בין הרכיבים.

### **8.2. המלצות מרכזיות לשלבי יישום**

פיתוח מערכת כה מורכבת מחייב גישה מדורגת:

* **שלב 1 (יסודות ובסיס ייחוס):** התמקדות בצנרת נתונים אמינה, חישוב מדויק של אינדיקטורים/LVN/MMD, ופיתוח בסיס ייחוס (baseline) עם סוכן RL יחיד עבור אסטרטגיית הליבה.1 שלב זה חיוני לאימות היתכנות האסטרטגיה הבסיסית לפני הוספת מורכבות MARL.
* **שלב 2 (ליבת MARL):** מימוש מבנה הסוכנים המתמחים (ATF30, ATF5, AMS), רשתות ההטמעה הייעודיות שלהם, ומסגרת CTDE/MAPPO עם מדיניות משותפת.
* **שלב 3 (מאפיינים מתקדמים):** שילוב סוכן הסיכון ההרמוני (AHR) ומנגנון אגרגציית ההחלטות מבוסס הקשב ("סופרפוזיציה").
* **שלב 4 (בדיקות קפדניות ופריסה):** ביצוע WFO מקיף, אופטימיזציית היפר-פרמטרים (HPO) באמצעות Optuna, שילוב HITL, ופריסה לסביבת מסחר על הנייר (paper trading) ולאחר מכן למסחר חי.

### **8.3. מענה לשאלות המחקר**

* **ארכיטקטורת רשת הטמעה:** ההמלצה היא להשתמש ב**רשתות הטמעה נפרדות ומותאמות** לכל סוג סוכן, בשל ההטרוגניות של נתוני הקלט שלהם. יש לאמן רשתות אלו מקצה-לקצה עם מערכת ה-RL. עבור סוכנים עם קלט דומה במבנהו (כמו ATF30 ו-ATF5), ניתן לשקול שימוש בתבנית ארכיטקטונית משותפת אך עם משקולות נפרדות.
* **היתכנות שיתוף מידע בין סוכנים בזמן אמת:** ההמלצה היא **להסתמך בשלב ראשון על תיאום מרומז** דרך פרדיגמת CTDE/MAPPO, מצב השוק המשותף (MMD), ומיזוג האותות על ידי סוכן הסיכון ההרמוני (AHR). יש להימנע מתקשורת ישירה בין שחקני הסוכנים אלא אם יוכח צורך מובהק, כדי לשמור על פשטות ולהימנע מצווארי בקבוק.
* **ארכיטקטורת פריסה ומנועי הסקה:** למסחר בנכס יחיד עם דרישות השהיה נמוכות, **Windows VPS הממוקם בקרבה פיזית לשרתי הבורסה** מועדף. יש למנף מנועי הסקה מהירים כגון **ONNX ו-TensorRT**. שילוב **HITL** חיוני לבקרה, ו-**Optuna** ישמש לאופטימיזציית היפר-פרמטרים.

### **8.4. מחקר עתידי והתפתחות המערכת**

* למידה דינמית של שבעת מצבי השוק על ידי מערכת ה-MMD, אם היא מתוכננת להיות יותר מסיווג סטטיסטי.
* בחינת פרוטוקולי תקשורת מתוחכמים יותר בין סוכנים אם תיאום מרומז אינו מספק.17
* פיתוח מנגנון ה"סופרפוזיציה" מעבר לאגרגציית קשב, למשל, בחינת שימוש באנסמבלים של מדיניויות משותפות.
* ניהול סיכונים אדפטיבי מעבר לקביעת גודל פוזיציה ו-SL/TP, למשל, למידה דינמית של התאמת קריטריוני ה"סינרגיה" עצמם.
* הרחבת המערכת למספר נכסים, מה שיגדיל משמעותית את המורכבות אך גם את הפוטנציאל של גישת ה-MARL.

## **9. מסקנות**

הדוח הציג תוכנית שאפתנית ומקיפה לפיתוח מערכת מסחר אלגוריתמית מרובת סוכנים (MARL) המיועדת למסחר בחוזים עתידיים. המערכת משלבת אסטרטגיית מסחר ייחודית עם טכנולוגיות חזיתיות בתחום הבינה המלאכותית, במטרה להשיג ביצועים אדפטיביים ועמידים. הדגש הושם על התאמת התכנון לדרישות המפורטות של המשתמש, כולל המעבר לנתוני ברים, הגדרת הסינרגיות והצורות ההרמוניות, והשימוש במנגנונים מתקדמים לזיהוי מצבי שוק ואגרגציית החלטות.

הפוטנציאל של מערכת כזו טמון ביכולתה לעבד מידע שוק מורכב באופן מבוזר, תוך שמירה על קוהרנטיות אסטרטגית גלובלית. זיהוי מצבי השוק באמצעות MMD נועד לאפשר למערכת להסתגל לתנאי שוק משתנים, בעוד שרעיון ה"סופרפוזיציה" שואף להקנות לה עמידות נוספת בפני רעש ואי-ודאות. אם תמומש בהצלחה, מערכת זו עשויה להציע ביצועים עדיפים על פני גישות מסחר מסורתיות או מודלים של סוכן יחיד, במיוחד בסביבות שוק דינמיות ותנודתיות.

עם זאת, מימוש המערכת כרוך בהתמודדות עם אתגרים משמעותיים. אלו כוללים את המורכבות הטכנית של כל רכיב בנפרד (MMD מבוסס חתימות, MARL עם מדיניות משותפת ותצפיות חלקיות, הנדסת מאפיינים מורכבת לצורות הרמוניות ועוצמת LVN, ומימוש פרקטי של "סופרפוזיציה" בהשראת קוונטים), ואת הקושי שבשילובם למערכת אחת קוהרנטית ויעילה. חלוקת הנתונים למטריצות Mi​ עבור הסוכנים השונים היא נקודה קריטית הדורשת איזון עדין. הבטחת "שיתוף פעולה מושלם" בין סוכנים הפועלים על סמך תצפיות חלקיות היא אתגר נוסף, וכך גם עיצוב פונקציית תגמול שתאזן בין יעדי רווח והפסד, ניהול סיכונים, והימנעות מהתאמת יתר.

למרות האתגרים, הרעיונות המוצגים במערכת נשענים במידה רבה על מחקרים וכלים קיימים בתחומי ה-MARL, ניתוח סדרות עתיות פיננסיות, וטכניקות למידת מכונה 17, מה שמגביר את היתכנות הפרויקט. הצלחת המימוש תהיה תלויה בגישה איטרטיבית לפיתוח, בבדיקות קפדניות בכל שלב, ובשילוב מנגנוני בקרה אנושיים (HITL) לתפעול אחראי ובטוח, במיוחד בהינתן הכוונה לסחור בכסף אמיתי. מערכת זו, אם תפותח ותאומת כראוי, עשויה להוות תרומה משמעותית לקידום השימוש ב-MARL במסחר אלגוריתמי מתוחכם.

#### Works cited

1. מחקר ממוקד לפיתוח מערכת מסחר אלגוריתמית מרובת סוכנים (MARL) מבוססת אסטרטגיה ספציפית V2
2. Unlocking Profits with the Machine Learning Momentum Indicator - Toolify.ai, accessed on June 11, 2025, <https://www.toolify.ai/ai-news/unlocking-profits-with-the-machine-learning-momentum-indicator-2641077>
3. Nadaraya-Watson Kernel Regression Using C# -- Visual Studio Magazine, accessed on June 11, 2025, <https://visualstudiomagazine.com/articles/2025/04/18/nadaraya-watson-kernel-regression-using-c.aspx>
4. MarketProfile - PyPI, accessed on June 11, 2025, <https://pypi.org/project/MarketProfile/>
5. bfolkens/py-market-profile: A library to calculate Market Profile (aka Volume Profile) for financial data from a Pandas DataFrame. - GitHub, accessed on June 11, 2025, <https://github.com/bfolkens/py-market-profile>
6. Futures Market Data - Live and historical futures data API - Databento, accessed on June 11, 2025, <https://databento.com/futures>
7. Real-Time Market Data - Low-latency APIs - Databento, accessed on June 11, 2025, <https://databento.com/live>
8. Market Data Feed Python API Description and Downloads - dxFeed, accessed on June 11, 2025, <https://dxfeed.com/api/python-api/>
9. Real-time and Delayed Data Service - dxFeed, accessed on June 11, 2025, <https://dxfeed.com/data-services/real-time-and-delayed-data-service/>
10. Triangle Price Pattern Detection In Python | Algorithmic Trading Indicator - YouTube, accessed on June 11, 2025, <https://www.youtube.com/watch?v=WVNB_6JRbl0>
11. Harmonic Pattern Recognition With Python - YouTube, accessed on June 11, 2025, <https://m.youtube.com/watch?v=ODHlC9YuowY&pp=ygUZI3B5dGhvbnBhdHRlcm5yZWNvZ25pdGlvbg%3D%3D>
12. Triangle Geometry Functions - 101 Computing, accessed on June 11, 2025, <https://www.101computing.net/triangle-geometry-functions/>
13. Triangle in python - geometry - Stack Overflow, accessed on June 11, 2025, <https://stackoverflow.com/questions/28596727/triangle-in-python>
14. Accessing the full scope of the higher timeframe datas on lower timeframes - Reddit, accessed on June 11, 2025, <https://www.reddit.com/r/TradingView/comments/1i80cik/accessing_the_full_scope_of_the_higher_timeframe/>
15. Higher-timeframe requests — Indicator by PineCoders - TradingView, accessed on June 11, 2025, <https://www.tradingview.com/script/W1YpYcOI-Higher-timeframe-requests/>
16. arXiv:2312.09009v2 [cs.AI] 3 Mar 2025, accessed on June 11, 2025, <https://arxiv.org/abs/2312.09009>
17. Learning Multi-Agent Intention-Aware Communication for Optimal ..., accessed on June 11, 2025, <https://seqml.github.io/marl4fin/>
18. Multi-Agent Hierarchical Graph Attention Actor–Critic Reinforcement ..., accessed on June 11, 2025, <https://www.mdpi.com/1099-4300/27/1/4>
19. Attention-Based Fault-Tolerant Approach for Multi-Agent Reinforcement Learning Systems - PMC - PubMed Central, accessed on June 11, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8469175/>
20. MARFT: Multi-Agent Reinforcement Fine-Tuning - arXiv, accessed on June 11, 2025, <https://arxiv.org/html/2504.16129v2>
21. Walk forward optimization - Wikipedia, accessed on June 11, 2025, <https://en.wikipedia.org/wiki/Walk_forward_optimization>
22. Walk Forward Optimization - QuantConnect.com, accessed on June 11, 2025, <https://www.quantconnect.com/docs/v2/writing-algorithms/optimization/walk-forward-optimization>
23. Best VPS for Algorithmic Trading in 2025 - XAUBOT, accessed on June 11, 2025, <https://xaubot.com/best-vps-for-algorithmic-trading/>
24. QuantVPS vs Beeks Proximity Cloud: Best VPS for Futures Trading in 2025, accessed on June 11, 2025, <https://www.quantvps.com/blog/quantvps-vs-beeks-proximity-cloud-best-vps-for-futures-trading-in-2025>
25. PyTorch + ONNX Runtime, accessed on June 11, 2025, <https://onnxruntime.ai/pytorch>
26. [D] Deploying Pytorch models. : r/MachineLearning - Reddit, accessed on June 11, 2025, <https://www.reddit.com/r/MachineLearning/comments/7ytxz4/d_deploying_pytorch_models/>
27. sithu31296/PyTorch-ONNX-TRT - TensorRT Conversion - GitHub, accessed on June 11, 2025, <https://github.com/sithu31296/PyTorch-ONNX-TRT>
28. [D] Looking for fastest inference way to run a pytorch model on TensorRT - Reddit, accessed on June 11, 2025, <https://www.reddit.com/r/MachineLearning/comments/1bmnn5j/d_looking_for_fastest_inference_way_to_run_a/>
29. What is Human-in-the-loop (HITL) in AI-assisted decision-making? - 1000minds, accessed on June 11, 2025, <https://www.1000minds.com/articles/human-in-the-loop>
30. Human-in-the-Loop: Balancing Automation and Expert Labelers, accessed on June 11, 2025, <https://keylabs.ai/blog/human-in-the-loop-balancing-automation-and-expert-labelers/>
31. Taming Communication and Sample Complexities in Decentralized Policy Evaluation for Cooperative Multi-Agent Reinforcement Learning, accessed on June 11, 2025, <https://par.nsf.gov/servlets/purl/10315977>
32. FinRL Contests: Benchmarking Data-driven Financial Reinforcement Learning Agents, accessed on June 11, 2025, <https://arxiv.org/html/2504.02281v3>
33. MultiAgentRL-trading/MARL - GitHub, accessed on June 11, 2025, <https://github.com/MultiAgentRL-trading/MARL>
34. [2501.06832] A novel multi-agent dynamic portfolio optimization learning system based on hierarchical deep reinforcement learning - arXiv, accessed on June 11, 2025, <https://arxiv.org/abs/2501.06832>
35. [2504.21048] Multi-Agent Reinforcement Learning for Resources Allocation Optimization: A Survey - arXiv, accessed on June 11, 2025, <https://www.arxiv.org/abs/2504.21048>