# פיתוח מערכת מסחר מרובת סוכנים מבוססת למידת חיזוק: אסטרטגיה, זיהוי משטרים, הטמעת נתונים ושיקולים מתקדמים

## I. מבוא: הצגת האתגר והחזון

היוזמה הנוכחית עוסקת בתכנון והקמה של מערכת מסחר אלגוריתמית מתוחכמת, המיועדת לפעול בשוק החוזים העתידיים. המערכת שואפת למזג באופן סינרגטי מספר טכנולוגיות מתקדמות מתחום הבינה המלאכותית ומדעי הנתונים, במטרה להשיג יתרון תחרותי בסביבת המסחר הדינמית. בליבת המערכת עומדת ארכיטקטורת למידת חיזוק מרובת סוכנים (MARL), אשר תוכננה לפעול בשיתוף פעולה הדוק.

ייחודה של המערכת טמון בשילוב של מספר מרכיבים מרכזיים:

1. **נתוני שוק והמרתם:** שימוש בנתוני נרות היסטוריים ונתונים בזמן אמת מספק הנתונים Rithmic, המכסים טווח של שנתיים לאחור באינטרוולים של 30 דקות ו-5 דקות. נרות אלו יומרו לנרות Heiken Ashi. המרה זו משרתת שתי מטרות עיקריות: ראשית, האסטרטגיה המיועדת פועלת באופן מיטבי יותר עם נרות Heiken Ashi; שנית, המרה זו מפחיתה את ממדיות הנתונים עבור כל נר, מה שעשוי להקל על החישובים של מנוע זיהוי המשטרים ועל עיבוד הנתונים על ידי הסוכנים. ככל שעולים בטווח הזמן, נר Heiken Ashi נוטה להציג תמונה ברורה ואינטרפרטטיבית יותר של המגמה בהשוואה לנרות יפניים רגילים.
2. **מנוע זיהוי משטרי שוק:** שילוב מנוע לזיהוי משטרי שוק, המבוסס על מדד Higher-Rank Maximum Mean Discrepancy (MMD) או חלופות יעילות, שיפעל על נרות Heiken Ashi בטווח זמן של 30 דקות. מנוע זה נועד לספק לסוכנים מודעות הקשרית רחבה יותר לגבי המצב המאקרו-כלכלי של השוק.
3. **אסטרטגיית מסחר מורכבת:** הטמעת אסטרטגיית מסחר ספציפית, הכוללת שילוב של אינדיקטורים כגון Machine Learning Momentum Index (MLMI), Nadaraya-Watson Rational Quadratic Kernel Regression (NW-RQK), ו-Fair Value Gaps (FVG), יחד עם ניתוח תחנות שוק (Market Profile LVNs) וצורות הרמוניות. אסטרטגיה זו, המכונה "פוליסי", תהווה את הבסיס לקבלת ההחלטות של הסוכנים.
4. **הטמעת נתונים באמצעות רשתות נוירונים:** שימוש ברשתות נוירונים (NN) להטמעת הנתונים הנכנסים לסוכנים (לאחר זיהוי משטר והמרת Heiken Ashi) לייצוג וקטורי דחוס. מטרת ההטבעה היא להפחית הטרוגניות וממדיות גבוהה של הנתונים, ולא לחלק את הנתונים למטריצות.
5. **ארכיטקטורת MARL:** פיתוח מערכת MARL שבה סוכנים מרובים פועלים עם מדיניות משותפת, אך כל אחד מהם עשוי לקבל "תצוגה" (view) שונה של הנתונים המוטבעים.
6. **סופרפוזיציה (בהשראת קוונטים):** חקירת שילוב עקרונות בהשראת סופרפוזיציה קוונטית כאמצעי להגברת העמידות והאמינות של המערכת.
7. **בדיקות מקיפות ומסחר חי:** המערכת מיועדת לפעול במסחר חי עם כסף אמיתי, ולכן נדרשת רובסטיות ודיוק מרביים, שיושגו באמצעות בדיקות לאחור (backtesting) קפדניות, אופטימיזציית Walk-Forward, ושימוש בסביבת Sandbox בזמן אמת.

האתגר המרכזי הוא לבנות מערכת שיודעת מה היא עושה "מהרגע הראשון עד לרגע האח लास्ट", כלומר, מערכת שיש לה הבנה עמוקה של כל שלבי התהליך, החל מאיסוף ועיבוד הנתונים, דרך קבלת החלטות מסחר מושכלות ועד לניהול סיכונים ובקרה. הדו"ח הנוכחי יפרט את תהליך הפיתוח, המבנה המוצע של הפרויקט, ויענה על שאלות ספציפיות הנוגעות לכל אחד ממרכיביו.

## II. תשתית נתונים ואסטרטגיית ליבה

תשתית נתונים איתנה ואסטרטגיית מסחר מוגדרת היטב מהוות את היסודות שעליהם תיבנה מערכת ה-MARL. שלב זה כולל את רכישת הנתונים, עיבודם הראשוני, והגדרת הלוגיקה הבסיסית של קבלת ההחלטות.

### א. רכישת נתונים, המרה ל-Heiken Ashi והכנתם

1. **מקור נתונים וטווחי זמן:** הנתונים ייאספו מספק הנתונים Rithmic. נתונים אלו יכללו היסטוריית מחירים של שנתיים אחורה, ברזולוציות זמן של 30 דקות ו-5 דקות. איכות הנתונים היא בעלת חשיבות עליונה, ויש להבטיח חותמות זמן מדויקות וטיפול עקבי בנתונים חסרים, אם קיימים.
2. **המרת נרות ל-Heiken Ashi:** כל נרות המחיר הגולמיים (יפניים סטנדרטיים) יומרו לנרות Heiken Ashi. המרה זו מבוצעת מכמה סיבות:
   * **התאמה לאסטרטגיה:** האסטרטגיה הספציפית שתיושם פועלת באופן יעיל יותר על נרות Heiken Ashi.
   * **הפחתת רעשים והחלקת מגמות:** נרות Heiken Ashi נוטים להחליק תנודות מחיר קצרות טווח ולהציג מגמות בצורה ברורה יותר. הדבר מקל על זיהוי כיוון השוק ומפחית "רעש" שעשוי להטעות את המודלים.
   * **הפחתת ממדים (במובן מסוים):** כל נר Heiken Ashi מחושב על בסיס מידע מהנר הקודם, מה שתורם להצגה "נקייה" יותר של תנועת המחיר. אף שאינה הפחתת ממדים במובן המתמטי הקלאסי (כמו PCA), היא מפשטת את ייצוג המידע פר-נר, מה שיכול להקל על חישובים עתידיים, כגון יצירת "נתיבי חתימה" (signature paths) עבור מנוע זיהוי המשטרים.
   * **שיפור אינטרפרטביליות:** ככל שעולים בטווחי הזמן, נר Heiken Ashi מציג תמונה פשוטה ואינטרפרטטיבית יותר של המגמה בהשוואה לנר רגיל.

הנוסחאות לחישוב נרות Heiken Ashi הן:

* + סגירה (HA-Close): (Open + High + Low + Close) / 4 (של הנר המקורי)
  + פתיחה (HA-Open): (HA-Open\_{previous} + HA-Close\_{previous}) / 2
  + גבוה (HA-High): \max(High\_{current}, HA-Open, HA-Close)
  + נמוך (HA-Low): \min(Low\_{current}, HA-Open, HA-Close)

1. **ויתור על נתוני רמה 2 (Level 2):** לאחר שיקול נוסף, הוחלט לוותר בשלב זה על שימוש בנתוני רמה 2 (עומק שוק). הסיבה המרכזית היא החשש מהכנסת "רעש" משמעותי למערכת. נתוני רמה 2 הם מורכבים, עתירי מידע, אך גם רגישים מאוד לתנודות קצרות טווח ולפעילות של אלגוריתמי מסחר בתדירות גבוהה (HFT). השילוב של חלוקת נתונים למטריצות והוספת מנוע זיהוי משטרים (שיתבסס על נרות 30 דקות) צפוי לספק מספיק מידע איכותי לסוכנים. הכנסת נתוני רמה 2, במיוחד לאור העובדה שהזיהוי יתקיים על טווחי זמן ארוכים יחסית (30 דקות לבר), עלולה להיות "אובדנית" מבחינת יחס אות לרעש. אף שניתן לשקול שימוש בטכניקות הפחתת ממדים כמו PCA/SVD על נתוני רמה 2, קיימת בעייתיות בנרמול נתונים הטרוגניים אלו באופן שישמר את המידע החשוב ולא יסיר בטעות ממדים קריטיים. לכן, ההתמקדות תהיה בנתוני נרות Heiken Ashi ובמידע ממנוע זיהוי המשטרים.

### ב. הגדרת אסטרטגיית המסחר ("פוליסי") המרכזית

אסטרטגיית המסחר ("פוליסי") היא לב המערכת והיא שתנחה את פעולות הסוכנים. האסטרטגיה מבוססת על שילוב של שלושה אינדיקטורים עיקריים הפועלים על טווחי זמן שונים, יחד עם ניתוח מבנה שוק נוסף.

1. **רכיבי האסטרטגיה:**
   * **Machine Learning Momentum Index (MLMI) (על גרף 30 דקות):** אינדיקטור מומנטום המשלב ממוצעים נעים של RSI ואלגוריתם k-Nearest Neighbors (k-NN) לחיזוי מומנטUM. האות המרכזי יהיה ככל הנראה חצייה בין קו החיזוי של ה-MLMI ל-WMA שלו.
   * **Nadaraya-Watson Rational Quadratic Kernel Regression (NW-RQK) (על גרף 30 דקות):** אינדיקטור רגרסיית גרעין לא-פרמטרית היוצר עקומת התאמה למחיר. האות יתקבל כאשר העקומה הופכת לעולה (שברי) או יורדת (דובי).
   * **Fair Value Gaps (FVG) - LuxAlgo (על גרף 5 דקות):** זיהוי "פערים" במחיר המייצגים חוסר איזון בין קונים למוכרים. האות הרלוונטי הוא "מיטיגציה" של FVG – חזרה של המחיר למילוי חלקי או מלא של הפער.
   * **ניתוח תחנות שוק (Market Profile LVNs) (על גרף 30 דקות):** זיהוי Low Volume Nodes (LVNs) באמצעות Market Profile. אזורים אלו מייצגים "תחנות" מחיר שבהן הייתה פעילות דלילה, ולעיתים קרובות המחיר נוטה "לרוץ" דרכן או להיבדק מחדש כרמות תמיכה/התנגדות. עוצמת התחנה תוערך על סמך התנהגות המחיר ההיסטורית סביבה.
   * **צורות הרמוניות (על גרף 5 דקות):** כאשר שלושת האותות מהאינדיקטורים (MLMI, NW-RQK, FVG) מתקבלים, הם יוצרים נקודות ציון. ניתן לראות את הקשר בין נקודות אלו כצורה גיאומטרית (למשל, משולש). תכונות הצורה (זוויות, יחסי צלעות) יכולות לספק מידע נוסף על איכות האות ולסייע בניהול הסיכונים.
2. **לוגיקת הכניסה ו"ארבע הסינרגיות":** כניסה לעסקה תתבצע רק כאשר *כל שלושת האינדיקטורים* (MLMI, NW-RQK, FVG mitigation) מספקים אות באותו כיוון. חשוב להבהיר: "ארבע הסינרגיות" אינן ארבע אסטרטגיות שונות שהסוכן בוחר ביניהן. במקום זאת, הן מייצגות את ארבעת הסדרים האפשריים שבהם שלושת האותות יכולים להופיע כדי ליצור הזדמנות כניסה תקפה. השוק הוא זה ש"בוחר" את הסינרגיה (כלומר, את סדר הופעת האותות). לדוגמה, עבור כניסה ללונג:
   * סינרגיה 1: אות MLMI (30 דק') ← אות FVG Mitigation (5 דק') ← אות NW-RQK (30 דק')
   * סינרגיה 2: אות MLMI (30 דק') ← אות NW-RQK (30 דק') ← אות FVG Mitigation (5 דק')
   * סינרגIA 3: אות NW-RQK (30 דק') ← אות FVG Mitigation (5 דק') ← אות MLMI (30 דק')
   * סינרגיה 4: אות NW-RQK (30 דק') ← אות MLMI (30 דק') ← אות FVG Mitigation (5 דק') (ובאופן דומה לשורט).

הסוכן לא יצטרך "לבחור" סינרגיה. תפקידו יהיה לזהות מתי אחת מארבע הסינרגיות הללו התממשה, ואז להחליט על אופן הכניסה וניהול הסיכונים, תוך התחשבות במידע הנוסף מצורות הרמוניות ומתחנות השוק (LVNs). לדוגמה, כניסה באזור LVN חזק, עם צורה הרמונית "איכותית", עשויה להצדיק גודל פוזיציה גדול יותר.

1. **מרחב הפעולות של הסוכן:** בהינתן האסטרטגיה, מרחב הפעולות של הסוכן יכלול:
   * **החלטת כניסה:** כן/לא (בהינתן התממשות סינרגיה).
   * **כיוון:** לונג/שורט (נקבע על ידי כיוון האותות).
   * **גודל פוזיציה:** מספר החוזים, שעשוי להיות מושפע מאיכות האות (למשל, עוצמת תחנת LVN, תכונות הצורה ההרמונית, ופרמטרי משטר השוק).
   * **ניהול עסקה:** הצבת פקודות הגנה (stop-loss) ויעד רווח (take-profit), שעשויים גם הם להיות דינמיים ותלויי הקשר. האסטרטגיה מנצלת את ה"פיגור" (lag) האינהרנטי של האינדיקטורים כיתרון, במטרה להיכנס למגמות לאחר שהן כבר החלו להתבסס (כ-10-15% לאחר תחילתן), מה שמאפשר הצבת פקודות הגנה הדוקות יותר. אין כוונה לייעל את פרמטרי האינדיקטורים ל"תוצאות על", אלא לשמור על יציבותם כברירת מחדל כדי להבטיח רובסטיות לאורך זמן ובשווקים שונים, במיוחד בשילוב עם מנוע זיהוי המשטרים וסוכן ה-RL.

### ג. מנוע זיהוי משטרי שוק (MMD)

זיהוי משטרי שוק הוא רכיב קריטי, שכן ביצועי אסטרטגיות עשויים להשתנות משמעותית בין משטרים שונים (למשל, שוק שורי, דובי, תנודתי, מגמתי).

1. **עקרונות MMD ו-MMD מסדר גבוה:** Maximum Mean Discrepancy (MMD) הוא מבחן סטטיסטי לא-פרמטרי המשמש להערכת הדמיון בין שתי התפלגויות הסתברות. הרעיון הוא להטמיע את ההתפלגויות במרחב מאפיינים בעל ממד גבוה (RKHS) ולמדוד את המרחק בין הייצוגים הממוצעים שלהן. בהקשר שלנו, MMD ישמש להשוואת התפלגות נתוני השוק (נרות Heiken Ashi של 30 דקות) בחלון זמן נוכחי להתפלגות בחלון זמן היסטורי. שינוי משמעותי יצביע על מעבר למשטר חדש. הגישה של "MMD מסדר גבוה" (Higher-Rank MMD), כפי שמוצע במחקר של Horvath ו-Issa (2023) ובמאגר הקוד issaz/signature-regime-detection, משתמשת בחתימות נתיבים מחוספסים (rough path signatures) כמפת מאפיינים. חתימות אלו לוכדות מידע על סדר התצפיות והתלות ביניהן לאורך זמן, ומתאימות לניתוח סדרות עתיות פיננסיות. היתרון הוא היכולת לשלב מידע פילטרציה (התפתחות המידע לאורך זמן). המטרה היא זיהוי שינויים התפלגותיים בזמן אמת. נבדקו מספר מאגרי קוד רלוונטיים:
   * tgcsaba/KSig: כלי לחישוב חתימות וגרעיני חתימה.
   * mcindoe/regimedetection: מכיל שיטות שונות לזיהוי משטרים, כולל כאלו המבוססות על שינויים סטטיסטיים.
   * luchungi/Generative-Model-Signature-MMD (והמזלג Afeks214/Generative-Model-Signature-MMD יחד עם המאמר arXiv:2407.19848v3): מציג שיטה המממשת חלקית את המתודולוגיה הרצויה. המחקר בוחן שימוש ב-PPO (Proximal Policy Optimization) בהקשר זה. השימוש ב-MAPPO (Multi-Agent PPO) בהקשר זה נראה פחות רלוונטי בשלב זה של זיהוי משטרים, שכן מדובר במשימה של ניתוח סדרת זמן יחידה (נתוני השוק) ולא באינטראקציה מרובת סוכנים.
2. **פרמטרים לסינון מצבי שוק:** הפלט של מנוע ה-MMD יכול להיות מדד מרחק או תווית קטגוריאלית של משטר השוק. פרמטרים אלו, או מאפיינים הנגזרים מהם (כגון רמת תנודתיות צפויה, עוצמת מגמה), ישולבו כמאפיינים נוספים עבור סוכני ה-RL. לאחר המרת הנרות ל-Heiken Ashi, יהיה קל יותר חישובית לייצר נתיבי חתימה (signature paths) ולנתח אותם לעומק כדי להבין את ההסתברויות ביניהם. קיימת שאלה האם הורדת הממד באמצעות Heiken Ashi מקלה או מקשה על המחקר. מצד אחד, היא מפחיתה רעש ומקלה על חישוב החתימות. מצד שני, היא עלולה להסיר מידע שעשוי להיות חיוני לזיהוי משטרים עדין. יש לבחון זאת אמפירית.
3. **שילוב פרמטרי MMD כהיפר-פרמטרים ל-Optuna:** נשאלת השאלה האם ניתן לקבוע שני פרמטרים ממנוע זיהוי המשטר (למשל, סף רגישות לשינוי משטר, או אורך חלון ההשוואה) ולהפוך אותם להיפר-פרמטרים ש-Optuna תכייל בזמן אמת (או במהלך אופטימיזציית היפר-פרמטרים).
   * **פוטנציאל:** כיול פרמטרים אלו עשוי לשפר את יכולת המערכת להסתגל לדינמיקות שוק משתנות בזמן אמת (אם הכיול הוא אונליין, מה שפחות סביר) או לשפר את ביצועי המודל באופן כללי (אם הכיול הוא אופליין כחלק מ-HPO). הדבר יכול לעזור להבין טוב יותר את הנתונים בזמן אמת ולשפר את הביצועים הסופיים.
   * **מורכבות:** כיול פרמטרים של מנוע זיהוי משטרים באמצעות Optuna הוא אפשרי, אך דורש הגדרה זהירה של פונקציית המטרה. אם מנוע ה-MMD הוא רכיב נפרד מה-RL, ייתכן שיהיה צורך לאמן/לכייל אותו בנפרד או בתהליך HPO מקונן. אם פרמטרי ה-MMD משפיעים ישירות על תצפית הסוכן, Optuna יכולה לנסות לכייל אותם כחלק מכיול כולל של המערכת.
   * **הסתייגות:** אם כבר קיימת רשת נוירונים שמחלקת נתונים בזמן אמת (כפי שנשקל בתחילה, אך הרעיון עודכן ל-NN להטמעה), הצורך בכיול היפר-פרמטרים של מנוע זיהוי משטרים עשוי להיראות פחות דחוף, אך כפי שהובהר, ה-NN ישמש להטמעה ולא לחלוקה. לכן, כיול פרמטרי ה-MMD נותר רלוונטי. הנושא יישאר למחקר נוסף כדי לקבוע את הגישה האופטימלית.
4. **חלופות ל-MMD:** יש לחקור האם קיימות שיטות יעילות יותר לזיהוי מצבי שוק, שאולי מהירות יותר או דורשות פחות משאבי חישוב מ-MMD מבוסס חתימות, מבלי להתפשר משמעותית על איכות הזיהוי.

## III. ארכיטקטורת מערכת MARL

לאחר עיבוד הנתונים והגדרת אסטרטגיית הליבה ומנוע זיהוי המשטרים, השלב הבא הוא תכנון ארכיטקטורת למידת החיזוק מרובת הסוכנים (MARL).

### א. הטמעת נתונים באמצעות רשתות נוירונים (NN Embedding)

1. **מטרת הרשת הנוירונית:** הובהר כי השימוש ברשת נוירונים (NN) אינו מיועד לחלוקת הנתונים למטריצות בזמן אמת. גישה כזו עלולה להוביל להתאמת יתר (overfitting) משמעותית, במיוחד עם נתונים לא-סטציונריים בזמן אמת, ולהוסיף סיבוכיות פיתוחית ורעש מיותר למערכת. במקום זאת, תפקיד ה-NN יהיה **להטמיע (embed)** את הנתונים הנכנסים לכל סוכן (לאחר המרת Heiken Ashi, הוספת פלט מנוע ה-MMD, ונתוני האסטרטגיה כמו LVNs וצורות הרמוניות) **לווקטור ייצוג דחוס ואינפורמטיבי**. מטרות ההטמעה הן:
   * **הפחתת הטרוגניות:** נתוני הקלט עשויים להיות ממקורות שונים ובסקלות שונות. ה-NN ילמד להמירם לייצוג אחיד.
   * **הפחתת ממדיות גבוהה:** אם וקטור המאפיינים הגולמי גדול, ה-NN ילמד ייצוג בממד נמוך יותר תוך שמירה על המידע הרלוונטי.
   * **יצירת סביבת אימון יציבה ואיכותית יותר לסוכנים:** וקטור מוטבע היטב יכול להקל על למידת המדיניות על ידי הסוכן. שיטות הטמעה יכולות לכלול רשתות קונבולוציה (CNN) אם יש מבנה מרחבי בנתונים (פחות סביר כאן), רשתות רקורנטיות (RNN/LSTM) אם יש חשיבות לסדר זמנים בתוך התצפית הבודדת של הסוכן, או רשתות Feedforward פשוטות (MLP) על וקטור מאפיינים שטוח.
2. **אימון הרשת הנוירונית עם הסוכן:** הרשת הנוירונית להטמעה אינה מאומנת בנפרד ואז "מוקפאת". היא מהווה חלק אינטגרלי מארכיטקטורת הסוכן עצמו.
   * **ארכיטקטורה משולבת:** ניתן לחשוב על זה כך: קלט הסוכן (נתוני שוק + MMD + אסטרטגיה) נכנס תחילה ל-NN Embedder. הפלט של ה-NN Embedder (וקטור ההטמעה) הופך להיות ה"מצב" (state) שהרשתות של סוכן ה-RL (רשת המדיניות ורשת הערך, אם קיימת) מקבלות כקלט.
   * **אימון מקצה לקצה (End-to-End):** כל המערך הזה (NN Embedder + רשתות RL) מאומן יחד, מקצה לקצה. הגרדיאנטים מאובדן ה-RL (למשל, מ-policy gradient או TD error) זורמים אחורה דרך רשתות ה-RL וגם דרך ה-NN Embedder.
   * **כיצד הסוכן "מאמת" את הרשת?** הסוכן אינו "מאמת" את הרשת במובן של בדיקה נפרדת. הלמידה המשותפת גורמת ל-NN Embedder ללמוד לייצר הטמעות שמועילות לסוכן ה-RL למקסם את התגמול. אם ההטמעות אינן טובות, הסוכן לא יצליח ללמוד מדיניות טובה, והאות לשיפור יתפשט גם ל-NN Embedder.
   * **מיקום הרשת:** הרשת להטמעה אינה "יושבת מעל הסוכנים". היא חלק מכל סוכן אינדיבידואלי. כל סוכן יקבל את התצפית הגולמית שלו, יעביר אותה דרך ה-NN Embedder (שעשוי להיות בעל משקולות משותפות לכל הסוכנים, או משקולות נפרדות אם הסוכנים הטרוגניים מאוד במהותם, אם כי כאן המדיניות משותפת), ואז יפעל על סמך ההטמעה.

הרעיון של הפשטת תצפיות באמצעות ייצוג מובנה, כפי שהוצע במחקר MARC , מדגיש את החשיבות של יצירת ייצוג איכותי מהתצפית הגולמית, וה-NN Embedder משרת מטרה זו.

### ב. חלוקת נתונים למטריצות (תצוגות סוכן)

1. **הגישה המקורית והשאלות שעלו:** הכוונה הראשונית הייתה לחלק את הנתונים באופן אקראי למטריצות M\_a, M\_b,..., M\_n, כאשר כל מטריצה מייצגת מקטע קטן יותר של הנתונים, וכל סוכן RL יופנה לכל מטריצה (או למטריצה ספציפית). עלתה השאלה מדוע לא כל מטריצה יכולה לכלול את כל סוגי המאפיינים, שכן ביום מסחר אחד צפויים לחוות מגוון מצבים. כמו כן, עלתה השאלה כיצד להבטיח שהמטריצות מתעדכנות בזמן אמת עבור מסחר חי.
2. **העדפה לחלוקה דינמית מבוססת משטר שוק:** במקום חלוקה אקראית סטטית, קיימת העדפה לחלוקה דינמית של הנתונים. הרעיון הוא שמנוע זיהוי המשטרים (MMD) יזהה את מצב השוק, ועל סמך מצב זה, הנתונים יחולקו או יסוננו באופן שיטייב את המידע עבור הסוכנים.
   * **מדוע דינמי?** חלוקה סטטית עלולה "לתקוע" את הסוכנים עם תצוגות שאינן רלוונטיות למצב השוק הנוכחי ולא תאפשר אדפטיביות. חלוקה דינמית יכולה להתאים את "תשומת הלב" של הסוכנים לחלקים החשובים ביותר של המידע בהתאם למשטר.
   * **כיצד זה יעבוד?** נדרשת מערכת שתקבל את הנתונים הגולמיים ואת פלט מנוע ה-MMD, ותבצע את החלוקה/סינון הדינמי. לדוגמה, במשטר מגמתי חזק, ייתכן שהסוכנים יקבלו תצוגות המתמקדות במאפייני מומנטום, בעוד שבמשטר תנודתי, התצוגות עשויות להתמקד במאפייני תנודתיות או חריגה מהממוצע.
   * **מ"מטריצות" ל"תצוגות מוטבעות":** לאחר החלוקה/סינון הדינמי, כל "תצוגה" (M\_i) שנוצרת עבור סוכן i תוזן ל-NN Embedder של אותו סוכן, כפי שתואר בסעיף הקודם. כך, כל סוכן מקבל וקטור הטמעה המותאם לתצוגה הספציפית שלו.
3. **אתגר שיתוף הפעולה עם תצפיות חלקיות:** כאשר סוכנים פועלים על סמך תצפיות חלקיות (גם אם דינמיות), האתגר הוא להבטיח שפעולותיהם, הנגזרות מאותה מדיניות משותפת, יהיו מתואמות ולא סותרות ברמה הגלובלית. המדיניות המשותפת צריכה להיות מסוגלת להתמודד עם מגוון רחב של תשומות מהתצוגות השונות. פלט מנוע ה-MMD, המספק הקשר גלובלי, הוא קריטי כאן. ייתכן שיהיה צורך במנגנוני תיאום נוספים או בעיצוב תגמול שיעניש חוסר תיאום. הצלחת המערכת תלויה ביכולת להבטיח שהסכום של הפעולות הלוקאליות-אופטימליות (הנגזרות מהמדיניות המשותפת על נתונים מקומיים) אכן יתקרב לאסטרטגיה גלובלית-אופטימלית.

### ג. ארכיטקטורת MARL מרכזית (CTDE) ומדיניות משותפת

1. **פרדיגמת אימון מרכזי עם ביצוע מבוזר (CTDE):** בהינתן הדרישה ל"שיתוף פעולה מושלם" ולמדיניות מסחר משותפת, פרדיגמת CTDE (Centralized Training with Decentralized Execution) נראית מתאימה. במהלך האימון, ניתן להשתמש במידע גלובלי (כמו מצבים ופעולות של כל הסוכנים, או פונקציית תגמול גלובלית) כדי לאמן את המדיניות המשותפת או פונקציית ערך מרכזית. במהלך הביצוע (מסחר חי), כל סוכן פועל באופן מבוזר על סמך התצפית המקומית המוטבעת שלו והמדיניות המשותפת.
2. **מדיניות משותפת ואסטרטגיית המסחר:** אסטרטגיית המסחר המפורטת (MLMI, NW-RQK, FVG, LVNs, Harmonics) תהווה את ליבת המדיניות (\pi) של כל סוכני ה-RL. כפי שהובהר, "ארבע הסינרגיות" הן דרכים שונות שבהן אותות הכניסה יכולים להתממש, והשוק הוא שקובע איזו מהן פעילה. תפקיד המדיניות המשותפת, אותה לומדים הסוכנים, הוא להחליט כיצד לפעול (גודל פוזיציה, ניהול סיכונים) כאשר סינרגיה כזו מתרחשת, בהתחשב בתצפית המקומית (המוטבעת) ובמצב השוק הגלובלי (ממנוע ה-MMD).
3. **ייצוג מצב ותצפיות ב-RLlib:** בספריות MARL כמו Ray RLlib, תצפיות מוחזרות בדרך כלל כמילונים הממפים מזהי סוכנים לתצפיות האישיות שלהם. מבנה זה מתאים למערכת המוצעת: כל סוכן i יקבל את התצפית שלו M\_i (לאחר חלוקה דינמית), יעביר אותה דרך ה-NN Embedder שלו, והווקטור המוטבע שיתקבל יהווה את המצב s\_i עבור אותו סוכן. המדיניות המשותפת \pi(a\_i | s\_i) תקבל את s\_i ותפיק פעולה a\_i.

### ד. שילוב סופרפוזיציה בהשראת קוונטים

הרעיון של שילוב "סופרפוזיציה" בהשראת מכניקת הקוונטים נועד להגביר את העמידות והאמינות של סוכני המסחר. הכוונה אינה לבנות מחשב קוונטי, אלא לשאוב השראה מרעיונות קוונטיים ולממשם באלגוריתמים קלאסיים.

1. **פענוח "סופרפוזיציה" בהקשר קלאסי:** בפיזיקה קוונטית, סופרפוזיציה מאפשרת למערכת להתקיים במספר מצבים בו-זמנית עד למדידה. באנלוגיה קלאסית, ניתן לחשוב על כך כ:
   * **אנסמבל של מדיניויות:** הפעלת מספר גרסאות של המדיניות המשותפת (אולי עם אתחולים שונים או שינויים קלים) במקביל. כל מדיניות באנסמבל מייצגת "מצב" בסופרפוזיציה. ההחלטה הסופית יכולה להתקבל על ידי הצבעה או ממוצע משוקלל.
   * **מדיניויות הסתברותיות:** המדיניות מפיקה התפלגות הסתברות על פני מרחב הפעולות, במקום פעולה דטרמיניסטית. הדבר מאפשר גמישות והימנעות מהתחייבות מוקדמת.
   * **מנגנון דמוי-QIDDM:** פעולות פוטנציאליות (הנגזרות מהסינרגיות) מקבלות הסתברויות. הסוכן נשאר ב"סופרפוזיציה" עד ש"טריגר קריסה" (אות חזק, שינוי משטר) מאלץ בחירה בפעולה ספציפית.
2. **השפעה על עמידות ואמינות:**
   * **עמידות מגיוון:** אם סוכן "שוקל" מספר השערות או אסטרטגיות בו-זמנית (למשל, דרך אנסמבל), הוא עמיד יותר לרעש או לשינויים פתאומיים. אם אסטרטגיה אחת נכשלת, אחרות עשויות להצליח.
   * **אמינות מהתחייבות מאוחרת:** "קריסה" לפעולה סופית רק לאחר איסוף ראיות מספיקות יכולה להוביל להחלטות אמינות יותר.
3. **מיקום בארכיטקטורה:** שילוב סופרפוזיציה יכול להתבצע ברמת הסוכן הבודד (כל סוכן מריץ אנסמבל פנימי או מדיניות הסתברותית) או ברמה מערכתית (אנסמבל של סוכנים). הדבר מגדיל מורכבות חישובית אך עשוי לשפר עמידות לאי-סטציונריות. ההחלטה על אופן המימוש המדויק תתקבל לאחר הבנה מעמיקה יותר של יכולות ומגבלות המערכת הבסיסית.הטבלה הבאה מסכמת גישות אפשריות לשילוב סופרפוזיציה:**טבלה 1: גישות סופרפוזיציה בהשראת קוונטים לעמידות סוכני RL** | גישה | רעיון "בהשראת קוונטים" מרכזי | מנגנון לעמידות/אמינות | תרחיש מימוש מעשי למסחר MARL | | :------------------------------------ | :------------------------------------------------------------- | :-------------------------------------------------------------------------------------------------------------------- | :----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------- | | ביצוע מדיניות באנסמבל | חקירה מקבילה של "מצבי" מדיניות מרובים. | גיוון אסטרטגי; אם מדיניות אחת נכשלת, אחרות עשויות להצליח. אגרגציית החלטות (הצבעה) יכולה לסנן רעשים. | כל סוכן (או המערכת כולה) מריץ K גרסאות של המדיניות המשותפת. החלטת המסחר הסופית מתקבלת בהצבעה או ממוצע משוקלל. | | בחירת פעולה הסתברותית | ייצוג מצב קוונטי כהתפלגות על פני מצבי בסיס. | הימנעות מהתחייבות מוקדמת לפעולה בודדת; מאפשר תגובה גמישה יותר לאי-ודאות. | רשת המדיניות מפיקה התפלגות הסתברות על פני מרחב הפעולות. הפעולה נדגמת או נבחרת הפעולה בעלת ההסתברות הגבוהה ביותר לאחר סף ביטחון. | | מעקב אחר השערות מצב מרובות | סוכן "קיים" במספר מצבי אמונה לגבי הסביבה. | יכולת להגיב למגוון רחב יותר של התפתחויות שוק אפשריות; פחות רגישות לזיהוי שגוי של המצב הנוכחי. | כל סוכן מחזיק התפלגות אמונה על פני מצבי שוק חבויים. המדיניות מותנית על התפלגות זו, או שכל השערה מזינה עותק של המדיניות, וההחלטות משוקללות. | | 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 סינרגיה. "קריסה" לסינרגיה ספציפית מתרחשת כאשר תנאי שוק מסוימים עוברים סף קריטי. |*מקור: עיבוד של , פרק 4.*

### ה. ספריות וכלים למימוש MARL

מימוש מערכת MARL מורכבת יסתמך על ספריות קיימות. Ray RLlib ו-PettingZoo הן רלוונטיות במיוחד.

* **Ray RLlib:** ספרייה ללמידת חיזוק התומכת בעומסים מבוזרים ובקנה מידה גדול, עם תכונות MARL חשובות.
* **PettingZoo:** תקן API לסביבות MARL, בדומה ל-Gymnasium. מאפשר יצירת סביבות מותאמות אישית. התכנון הסביר יכלול יצירת סביבת מסחר מותאמת אישית העומדת בתקן PettingZoo, ולאחר מכן שימוש ב-RLlib לאימון הסוכנים בסביבה זו.

## IV. אימון, הערכה ואופטימיזציה

לאחר תכנון הארכיטקטורה, יש לאמן את המערכת, להעריך את ביצועיה בקפדנות, ולבצע אופטימיזציה של הפרמטרים.

### א. הגדרת קריטריונים להצלחה ופונקציית תגמול גלובלית

1. **קריטריונים להצלחה מעבר לרווח והפסד (P&L):** התמקדות צרה ביעד P&L קצר טווח עלולה להוביל להתאמת יתר. יש לאמץ מגוון מדדי ביצועים :
   * **מדדי תשואה מותאמת לסיכון:** יחס שארפ, יחס סורטינו, יחס קלמאר.
   * **מדדי רווחיות:** תשואה מצטברת, פקטור רווח.
   * **מדדי סיכון:** משיכה מקסימלית (MDD), תנודתיות התשואות.
   * **מדדי יעילות מסחר:** אחוז עסקאות רווחיות, תוחלת רווח לעסקה. הצלחת המערכת תימדד גם ביכולתה לשמור על ביצועים עקביים על פני משטרי שוק שונים ומכשירים פיננסיים שונים.
2. **עיצוב פונקציית תגמול גלובלית (Global Reward Function):** פונקציית התגמול צריכה לשקף את מטרות הצוות הכוללות. היא תחושב באופן גלובלי עבור כל הסוכנים. הספרייה vectorbt.dev הוזכרה ככלי מועדף לביצוע בקטסטים מהירים ומדויקים, הודות לחישוב מקבילי באמצעות וקטורים. ניתן להשתמש בפלט של vectorbt (כגון P&L, יחס שארפ על תקופה) כדי *לחשב* את התגמול הגלובלי שיינתן לסוכנים.
   * **תפקיד vectorbt:** vectorbt אינה הסביבה האינטראקטיבית עצמה המספקת תגמולים צעד-אחר-צעד במהלך לולאת ה-RL. היא משמשת להערכת הביצועים של רצף פעולות של הסוכנים על פני תקופה (למשל, בסוף אפיזודת אימון). התוצאות מ-vectorbt (כמו רווח/הפסד כולל, יחס שארפ) יכולות לשמש כרכיב מרכזי בתגמול האפיזודי.
   * **רכיבי התגמול:** התגמול יכול לכלול שינוי בערך התיק, יחס שארפ על חלון אחרון, עונשים על סיכון מופרז או משיכות גדולות, ואף תגמול על ביצוע מוצלח של הסינרגיות האסטרטגיות בתנאים המתאימים.

### ב. הקצאת אשראי (Credit Assignment)

במערכת MARL שיתופית עם תגמול גלובלי, אתגר מרכזי הוא כיצד לקבוע את תרומתו של כל סוכן בודד להצלחה או לכישלון הכולל. אם כל הסוכנים חולקים את אותה מדיניות ומקבלים את אותו תגמול גלובלי, בעיית הקצאת האשראי הקלאסית (מי הסוכן ש"אשם" או "זכאי"?) פחות חריפה מאשר במערכות עם סוכנים הטרוגניים לחלוטין. עם זאת, אם התצוגות (M\_i) של הסוכנים שונות מספיק כדי להוביל להבדלים בפעולות המוצעות (לפני אגרגציה כלשהי, אם קיימת), או אם מנגנון הסופרפוזיציה כולל רכיבים שונים הפועלים במקביל, אזי הבנת התרומה של כל רכיב/תצוגה הופכת לרלוונטית. שיטות כמו Value Decomposition Networks (VDN) או Counterfactual Rewards עשויות להיות רלוונטיות אם יהיה צורך בניתוח תרומה מעמיק יותר, במיוחד אם המערכת תתפתח לכיוונים מורכבים יותר. בשלב ראשון, עם מדיניות משותפת ותגמול גלובלי, ההתמקדות תהיה בלמידת המדיניות האופטימלית המשותפת. הכוונה היא לבחון את נושא הקצאת האשראי לעומק, ואם יימצא כי גישה מסוימת טובה, לממש אותה גם במהלך האימון וגם במסחר חי.

### ג. אופטימיזציית היפר-פרמטרים (HPO)

מערכת MARL מורכבת כוללת היפר-פרמטרים רבים: שיעורי למידה, ארכיטקטורות רשת, פרמטרי מנוע MMD, פרמטרי מודול סופרפוזיציה, ואף פרמטרים אסטרטגיים (כמו מספר סוכנים/תצוגות, אופן חלוקת הנתונים אם הוא מכוונן).

* **אופן הפעולה של HPO:** ספריות כמו Optuna, בשילוב עם Ray Tune (המשולבת ב-RLlib), מאפשרות אוטומציה של תהליך זה. התהליך כולל הגדרת פונקציית "מטרה" (objective function) שעוטפת את אימון המודל והערכתו על סט ולידציה, ומחזירה מדד ביצועים (למשל, יחס שארפ). Optuna משתמשת באלגוריתמים (כמו אופטימיזציה בייסיאנית) לחיפוש יעיל במרחב ההיפר-פרמטרים כדי למקסם/למזער את פונקציית המטרה.
* **ייצוב המערכת:** כיול היפר-פרמטרים נכון הוא קריטי לייצוב המערכת ולהשגת ביצועים אופטימליים ועקביים. הוא מונע בחירות שרירותיות של פרמטרים שעשויות לעבוד טוב על סט נתונים אחד אך להיכשל על אחר.

### ד. רגולריזציה

רגולריזציה היא טכניקה למניעת התאמת יתר (overfitting) על ידי הוספת "עונש" למורכבות המודל.

* **בחירת סוג הרגולריזציה:** במודלים גדולים עם זרימה מתמשכת של דאטה, רגולריזציית L2 (המענישה משקולות גדולות ברשת הנוירונים) היא לרוב בחירה טובה. היא מעודדת את המודל להשתמש בכל המאפיינים במידה מסוימת, בניגוד ל-L1 שעשויה לאפס משקולות של מאפיינים מסוימים ובכך "להרוס רצפי אימון" או להתעלם ממידע שעשוי להיות שימושי בהקשרים מסוימים.
* **מיקום בארכיטקטורה:** רגולריזציית L2 (או L1) מיושמת בדרך כלל על שכבות המשקולות של הרשתות הנוירוניות (הן ה-NN Embedder והן רשתות המדיניות/ערך של סוכן ה-RL). רוב מסגרות הלמידה העמוקה (כמו TensorFlow, PyTorch) מאפשרות הוספת רגולריזציה לשכבות בקלות. ניתן גם להוסיף עונש על מורכבות המדיניות ישירות לפונקציית המטרה של אלגוריתם ה-RL.

## V. בדיקות, ולידציה ופריסה

תהליך הבדיקה והוולידציה הוא קריטי להבטחת אמינות המערכת לפני פריסה למסחר חי.

### א. אימון מקיף ובדיקות Walk-Forward (WF)

1. **מטרת אימון מקיף (Walk-Forward):** לאחר שהמערכת מוכנה ברמה הראשונית, יש לבצע אימון מקיף בשיטת Walk-Forward Optimization (WFO). WFO היא טכניקה קפדנית לבדיקת אסטרטגיות מסחר המדמה טוב יותר תנאי מסחר אמיתיים ומסייעת בהפחתת התאמת יתר. ב-WFO, הנתונים ההיסטוריים מחולקים לתקופות אימון (In-Sample, IS) ובדיקה (Out-of-Sample, OOS) מתגלגלות. המערכת מאומנת (או מכוילת היפר-פרמטרית) על תקופת IS, ולאחר מכן ביצועיה נבדקים על תקופת ה-OOS הבאה מיד אחריה, שאינה נראתה למודל במהלך האימון. התהליך חוזר על עצמו, כאשר חלון ה-IS/OOS "מתקדם" לאורך ציר הזמן.
2. **אופן הביצוע בפועל:**
   * **על מה עושים WF?** ה-WF יתבצע על המערכת הכוללת. זה אומר שכל לולאת ה-WF תכלול:
     + **שלב IS:** אימון סוכני ה-MARL (כולל ה-NN Embedder שלהם), תוך שימוש בנתוני ה-IS. אם מתבצעת גם אופטימיזציית היפר-פרמטרים (HPO) עם Optuna, היא תתבצע על נתוני ה-IS (עם חלוקה פנימית ל-train/validation בתוך ה-IS). מנוע ה-MMD גם הוא יפעל על נתוני ה-IS (או יכויל עליהם, אם פרמטריו דינמיים).
     + **שלב OOS:** לאחר שהמודלים אומנו/כוילו על ה-IS, הם "מוקפאים" ונבדקים על נתוני ה-OOS הבאים מיד. הביצועים על נתוני ה-OOS הם אלו שקובעים את איכות המערכת, שכן הם מדמים מסחר על נתונים "חדשים".
   * **שימוש ב-vectorbt ב-WF:** ספריית vectorbt תשמש בשלב ה-OOS של כל חלון WF. לאחר שהסוכנים מייצרים החלטות מסחר על נתוני ה-OOS, vectorbt תשמש כדי לחשב במהירות את מדדי הביצועים (P&L, Sharpe, MDD וכו') על תקופת OOS זו. מדדים אלו ייאספו מכל חלונות ה-OOS כדי לתת תמונה כוללת של ביצועי המערכת.
   * **פרמטרים לכוונון ב-WF/HPO:** יכללו את אלו של אלגוריתם ה-MARL (שיעורי למידה, ארכיטקטורות רשת), פרמטרי מנוע ה-MMD (אם הוחלט שהם דינמיים), פרמטרי מודול הסופרפוזיציה, ואף פרמטרים אסטרטגיים כמו אופן החלוקה הדינמית של הנתונים (אם יש פרמטרים שמגדירים אותה).

הטבלה הבאה מציגה דוגמה מבנית לתהליך Walk-Forward:**טבלה 2: מבנה תהליך Walk-Forward Optimization** | איטרציית WF | תקופת In-Sample (IS) | תקופת Out-of-Sample (OOS) | פוקוס אימון/כיול ב-IS | מדדי הערכה מרכזיים ב-OOS | | :---------- | :-------------------- | :------------------------- | :----------------------------------------------------------------------------------- | :----------------------- | | 1 | ינואר 2022 - דצמבר 2022 | ינואר 2023 - מרץ 2023 | אימון ראשוני של MARL, כיול HPO של רשתות, MMD, סופרפוזיציה. | Sharpe, MDD, P&L | | 2 | אפריל 2022 - מרץ 2023 | אפריל 2023 - יוני 2023 | אימון מחדש (fine-tuning) של MARL, עדכון כיול HPO על בסיס נתונים עדכניים יותר. | Sharpe, MDD, P&L | |... |... |... |... |... | | N | אוקטובר 2023 - ספטמבר 2024 | אוקטובר 2024 - דצמבר 2024 | אימון מחדש אחרון, כיול HPO סופי. | Sharpe, MDD, P&L | *ביצועי OOS מצטברים מכל האיטרציות יספקו הערכה רובסטית של המערכת.*

### ב. מסגרת בדיקות לאחור (Backtesting) רובסטית

1. **מעבר ל-vectorbt עבור סימולציית מערכת מלאה:** בעוד ש-vectorbt מצוינת לבקטסטים וקטוריים מהירים (הערכת לוגיקת אות-למסחר), אימון ה-MARL עצמו דורש *סביבה אינטראקטיבית*. סביבה זו מדמה אינטראקציות שוק: הסוכן מבצע פעולה, מצב הסביבה משתנה, וניתן תגמול. סביבה זו יכולה להשתמש בנתונים היסטוריים (בדומה ל-vectorbt) אך צריכה לספק תצפיות ותגמולים צעד-אחר-צעד עבור לולאת האימון של ה-RL (למשל, בהתאם ל-API של PettingZoo).
2. **פרטי סימולציה ריאליסטיים:** הבקטסט חייב לכלול עלויות עסקה (עמלות, החלקה), השפעה על השוק (אם רלוונטי), וטיפול בדינמיקת ספר פקודות אם משתמשים בפקודות גבול.

### ג. ארגז חול (Sandbox) בזמן אמת

1. **חשיבות ארגז החול:** ארגז חול בזמן אמת (או קרוב לזמן אמת) הוא קריטי לפני פריסה חיה. הוא מאפשר לבדוק את כלל המערכת עם נתונים חיים (או מושהים) מספק כמו Rithmic, לאמת חישוב תגמולים, לצפות בהתנהגות הסוכנים ולבדוק יציבות.
2. **תפקיד ארגז החול (ולא שינוי מודל בזמן אמת):** הרעיון להפחית רגולריזציה של מודל בזמן אמת על סמך ביצועי ארגז החול הוא מסוכן ולא מקובל. רגולריזציה נקבעת באימון לשיפור הכללה. שינויה הדינמי עלול להוביל לחוסר יציבות. ארגז החול משמש בעיקר *לוולידציה* של המודל המאומן ושל המערכת כולה בסביבה ריאליסטית. הוא מאפשר *לאסוף נתונים* על ביצועי המערכת הנוכחית. נתונים אלו יכולים לאחר מכן לשמש *לעדכון ושיפור המודל במחזור האימון הבא (אופליין)*, כולל כיול מחדש של רגולריזציה או לוגיקת הקצאת אשראי, אם יש צורך. הוא עוזר לזהות האם המודל הנוכחי מתנהג כמצופה והאם אותות התגמול משמעותיים, ולא לשינוי דינמי של מאפייני מודל יסודיים בזמן אמת.

### ד. שילוב אדם בלולאה (HITL)

1. **הכרחיות HITL:** שילוב אדם בלולאה (Human-in-the-Loop) הוא הכרחי במערכת מסחר עם כסף אמיתי, ומאפשר בקרה וניהול סיכונים.
2. **תפקידי HITL:**
   * אימות זיהויי משטר MMD.
   * פיקוח על פריסת האסטרטגיה ופעולות הסוכנים.
   * יכולת התערבות (כיבוי מערכת, שינוי פרמטרים) במצבי קיצון.
   * הנחיית עידון המדיניות על סמך ביצועים והבנת שוק. מעבר לעצירת חירום, HITL יכול לספק תובנות איכותיות (למשל, השפעת אירוע גיאופוליטי) שיכולות להנחות אימון מחדש או שיפור פונקציית התגמול.

### ה. אסטרטגיות אגרגציה להחלטות סוכנים

אם מספר סוכנים (או חברי אנסמבל מסופרפוזיציה) מציעים פעולות, יש לקבוע כיצד מתקבלת החלטת סחר נטו אחת.

* **צורך באגרגציה:** תלוי באופן מימוש הסופרפוזיציה ובמידת השונות בין תצוגות הסוכנים. אם יש סוכן "לוגי" אחד, או שכל הסוכנים צפויים להגיע לאותה החלטה, האגרגציה טריוויאלית.
* **שיטות אפשריות:** הצבעת רוב, ממוצע משוקלל של הסתברויות פעולה, או שיטות מורכבות יותר כמו VDN/QMIX המטפלות באגרגציה באופן אימפליציטי. אף שיש גמישות להחליט על כך בהמשך, חשוב לתכנן מראש אסטרטגיות אגרגציה ראשוניות.

## VI. מפת דרכים ליישום ומסלולי מחקר עתידיים

מימוש מערכת מורכבת זו דורש תוכנית שלבית וראייה לעתיד.

### א. תוכנית פיתוח שלבית

הטבלה הבאה מתווה מפת דרכים אפשרית ליישום המערכת:

**טבלה 3: מפת דרכים ליישום שלבי עם אבני דרך** | שלב # | שם השלב | אבני דרך / תוצרים מרכזיים | מורכבות/משך מוערך (איכותני) | טכנולוגיות/כלים מרכזיים | | :---- | :---------------------------------------- | :------------------------------------------------------------------------------------------------------------------------------------------------------ | :-------------------------- | :--------------------------------------------------------- | | 1 | תשתית נתונים ואסטרטגיית בסיס (Baseline) | צנרת נתונים מ-Rithmic, המרת Heiken Ashi. מימוש אסטרטגיית ליבה (MLMI, NW-RQK, FVG, LVN, הרמוניות) ב-vectorbt כבסיס לא-RL. ולידציית אותות. | בינונית | Python, Pandas, vectorbt, Rithmic API | | 2 | פיתוח וולידציית מנוע MMD | מימוש טכניקת MMD נבחרת (למשל, מבוססת חתימות או חלופה). בדיקה על נתונים היסטוריים. הגדרת פלט משטרים. | גבוהה | Python, iisignature (אופציונלי), ספריות MMD ייעודיות | | 3 | מערכת RL עם סוכן יחיד | פיתוח סביבת MARL בסיסית (PettingZoo API). שילוב MMD. מימוש NN Embedding. אימון סוכן RL יחיד (N=1) לביצוע אסטרטגיית הליבה. | גבוהה | PettingZoo, Ray RLlib, TensorFlow/PyTorch | | 4 | הרחבת MARL ושילוב סופרפוזיציה | אם N>1: מימוש חלוקת נתונים דינמית/תצוגות והקצאת אשראי. שילוב מנגנון סופרפוזיציה נבחר (למשל, אנסמבל). | גבוהה מאוד | Ray RLlib, TensorFlow/PyTorch | | 5 | בדיקות קפדניות, WFO, ו-HPO | ביצוע Walk-Forward Optimization. אופטימיזציית היפר-פרמטרים עם Optuna/Ray Tune. | גבוהה | vectorbt, Optuna, Ray Tune | | 6 | פריסה לארגז חול וניטור | פריסת המערכת המלאה לארגז חול בזמן אמת. הפעלת מנגנוני HITL. | בינונית-גבוהה | תשתית פריסה, כלי ניטור, Rithmic (חיבור חי) | | 7 | פריסה חיה ושיפור מתמיד | מסחר חי מבוקר. ניטור רציף, איסוף נתונים, ואיטרציות שיפור למודל ולאסטרטגיה. | מתמשך | כל הכלים הנ"ל |

*מקור: עיבוד של , פרק 6.1, והתאמות לדרישות המשתמש.*

חשיבות רבה נודעת לבניית בסיסי ייחוס (baselines) בכל שלב. ראשית, יש לבדוק את אסטרטגיית האותות הטהורה (ללא RL) באמצעות vectorbt. לאחר מכן, סוכן RL יחיד ינסה לשפר ביצועים אלו. לבסוף, מערכת ה-MARL המלאה או זו עם סופרפוזיציה תיבחן מול סוכן ה-RL היחיד. גישה זו מסייעת לבודד את הערך המוסף של כל שכבת מורכבות.

### ב. התמודדות עם אתגרי מערכת מרכזיים

* **מדרגיות (Scalability):** שימוש ב-Ray RLlib, טיפול יעיל בנתונים.
* **אי-סטציונריות (Non-stationarity):** מנוע MMD, WFO, יכולות למידה מתמשכת, סופרפוזיציה לאדפטיביות.
* **נצפות חלקית (Partial Observability) (אם N>1 עם M\_i שונים):** ייצוג מצב עשיר, הקשר גלובלי מ-MMD, תקשורת פוטנציאלית.
* **פרשנות/הסברתיות (XAI):** זהו אתגר במערכות MARL ו-NN. יש לשקול רישום מפורט, ויזואליזציה של החלטות סוכנים מול קלט, ומנגנוני תשומת לב (attention) אם משתמשים בטרנספורמרים ב-NN. חיוני לאמון מצד ה-HITL.

### ג. אפיקי מחקר עתידיים והתפתחות המערכת

1. **טכניקות MMD חלופיות/מהירות יותר:** חקירת שיטות זיהוי משטרים שאולי מהירות יותר או דורשות פחות משאבים מ-MMD מבוסס חתימות, מבלי להתפשר משמעותית על איכות הזיהוי.
2. **חלוקת נתונים דינמית/בחירת מאפיינים מבוססת מצב שוק:** מעבר מחלוקה סטטית למערכת שבה תצוגות הנתונים (M\_i) מותאמות דינמית למשטר השוק הנוכחי. זה יכול לכלול מנגנוני תשומת לב או בחירה מבוססת-חוקים של מאפיינים עבור M\_i.
3. **אבולוציה של סינרגיות:** חקירת האפשרות שהמערכת תלמד או תפתח באופן אוטונומי סינרגיות אסטרטגיות חדשות, מעבר לאלו שהוגדרו מראש. זוהי בעיית RL מתקדמת יותר (גילוי אסטרטגיה לעומת ביצוע אסטרטגיה).
4. **שילוב עמוק יותר של עקרונות בהשראת קוונטים:** מעבר לאנלוגיות קלאסיות נוכחיות, חקירת מחקרים חדשים יותר ב-QRL שעשויים להציע יתרונות עמוקים יותר (שעדיין ניתנים למימוש קלאסי).
5. **פרשנות משופרת:** מחקר ופיתוח שיטות XAI המותאמות ספציפית למערכות MARL פיננסיות.
6. **למידה מתמשכת ארוכת-טווח:** פיתוח ארכיטקטורות המאפשרות למערכת להסתגל באופן רציף לשינויים ארוכי-טווח בדינמיקת השוק, מעבר לזיהוי משטרים טקטי.

שאלות המחקר העתידיות מצביעות על חזון למערכת בעלת יכולת הסתגלות עצמית גבוהה, שלא רק מבצעת אסטרטגיה אלא גם משכללת באופן מתמיד את הבנתה ועיבודה של מידע שוק. הארכיטקטורה הראשונית צריכה להיות מודולרית מספיק כדי לאפשר חקירה ושילוב של יכולות מתקדמות אלו בהמשך.

## VII. מסקנות והנחיות לעתיד

המחקר הנוכחי הציג תוכנית מקיפה לפיתוח מערכת מסחר מרובת סוכנים (MARL) מתקדמת, המיועדת למסחר בחוזים עתידיים תוך שימוש בנתוני Heiken Ashi, מנוע זיהוי משטרי שוק (כגון MMD), אסטרטגיית מסחר מורכבת, הטמעת נתונים באמצעות רשתות נוירונים, ורעיונות בהשראת סופרפוזיציה קוונטית לשיפור עמידות המערכת.

**פוטנציאל המערכת ואתגריה המרכזיים:** הפוטנציאל טמון ביכולת לעבד מידע שוק מורכב באופן מבוזר, תוך שמירה על קוהרנטיות אסטרטגית והסתגלות לתנאי שוק משתנים. אם תמומש בהצלחה, מערכת זו עשויה להציע ביצועים עדיפים על פני גישות מסורתיות, במיוחד בשווקים דינמיים. עם זאת, המימוש כרוך בהתמודדות עם אתגרים משמעותיים: המורכבות הטכנית של כל רכיב, הקושי בשילובם, הבטחת שיתוף פעולה אפקטיבי בין סוכנים עם תצפיות חלקיות, ועיצוב פונקציית תגמול מאוזנת. ההחלטה להשתמש בנתוני Heiken Ashi ולהטמיעם באמצעות NN, תוך ויתור על נתוני רמה 2 בשלב זה, נועדה למקד את המאמץ ולהפחית רעשים.

**היתכנות על בסיס מחקר וכלים קיימים:** למרות האתגרים, הרעיונות המוצגים נשענים על מחקרים קיימים וכלים זמינים. ספריות כמו Ray RLlib ו-PettingZoo, יחד עם מחקרים בתחום זיהוי משטרים וטכניקות בהשראת קוונטים, מספקות בסיס איתן. המשימה שאפתנית אך ניתנת לביצוע הדרגתי.

**הנחיות מרכזיות ליישום:**

1. **פיתוח הדרגתי ומבוסס בסיסי ייחוס:** התקדמות שלבית, החל מתשתית נתונים, דרך אסטרטגיית בסיס, מנוע MMD, סוכן RL יחיד, ועד למערכת MARL מלאה עם סופרפוזיציה. כל שלב צריך להימדד מול קודמו.
2. **התמקדות באיכות הנתונים והטמעתם:** נתוני Heiken Ashi איכותיים והטמעה יעילה באמצעות NN הם קריטיים להצלחת הסוכנים.
3. **חלוקת נתונים דינמית:** יש לתעדף פיתוח מנגנון לחלוקת/סינון נתונים דינמי המבוסס על פלט מנוע זיהוי המשטרים, על פני חלוקה אקראית סטטית.
4. **אסטרטגיה ("פוליסי") מוגדרת היטב:** יש לתרגם את לוגיקת ה-MLMI, NW-RQK, FVG, LVNs, והצורות ההרמוניות לקוד רובסטי שישמש כמדיניות המשותפת.
5. **בדיקות קפדניות:** WFO, בקטסטינג עם vectorbt, וארגז חול בזמן אמת הם הכרחיים לפני מסחר חי.
6. **שילוב HITL:** לא רק כרשת ביטחון, אלא גם כמקור להנחיה ושיפור איכותני של המערכת.

**הנחיות למחקר עתידי:** הפיתוח והבדיקה של מערכת זו יניבו תובנות שיכולות להנחות מחקר עתידי בתחומים כמו טכניקות MMD מתקדמות, חלוקת נתונים לומדת, אבולוציה של אסטרטגיות, שילוב עמוק יותר של עקרונות בהשראת קוונטים, פרשנות, ולמידה מתמשכת ארוכת-טווח.

לסיכום, המערכת המוצעת מייצגת חזון שאפתני וחדשני למסחר אלגוריתמי. מימושה ידרוש מאמץ מחקרי ופיתוחי משמעותי, אך הפוטנציאל לקידום התחום ולהשגת ביצועי מסחר משופרים מצדיק את המאמץ. התקדמות הדרגתית, התמודדות שיטתית עם האתגרים, והתבססות על מחקר וכלים קיימים הם המפתח להפיכת חזון זה למציאות.